

Concurrent Band Selection and Traversability Estimation from Long-Wave Hyperspectral Imagery in Off-Road Settings

Florence Yellin¹, Scott McCloskey¹, Cole Hill², Eric Smith¹, and Brian Clipp¹

¹Kitware {florence.yellin, scott.mccloskey, eric.smith, brian.clipp}@kitware.com

²University of South Florida coleh@usf.edu

Abstract

Autonomous navigation has become increasingly popular in recent years; However, most existing methods focus on on-road navigation and utilize active sensors, such as LiDAR. This paper instead focuses on autonomous off-road navigation using traversability estimation from passive sensors, specifically long-wave (LW) hyperspectral imagery (HSI). We present a method for selecting a subset of hyperspectral bands that are most useful for traversability estimation by designing a band selection module that designs a minimal sensor that measures sparsely-sampled spectral bands while jointly training a semantic segmentation network for traversability estimation. The effectiveness of our method is demonstrated using our dataset of LW HSI from diverse off-road scenes including forest, desert, snow, ponds, and open fields. Our dataset includes imagery collected both during the daytime and nighttime during various weather conditions, including challenging scenes with a wide range of obstacles. Using our method, we learn a small subset (2%) of all the HSI bands that can achieve competitive or better traversability estimation accuracy to that achieved when utilizing all hyperspectral bands. Using only 5 bands, our method is able to achieve a mean class accuracy that is only 1.3% less than that achieved using full 256-band HSI and only 0.1% less than that achieved using 250-band HSI, demonstrating the success of our method.

1. Introduction

Autonomous, off-road navigation is important for a range of applications, from space exploration to terrestrial search and rescue. Recent focus on autonomous navigation has been mostly focused on *on-road* navigation, but methods designed for on-road conditions often fail when applied to *off-road* environments [6, 7, 27]. Furthermore, much of the existing literature uses active sensors, such as LiDAR, for autonomous navigation; however, for certain off-road applications, active sensors may not be available due to cost, weight, power, or other constraints.

In this paper, we demonstrate the feasibility of using passive sensors, specifically LW hyperspectral sensors, for

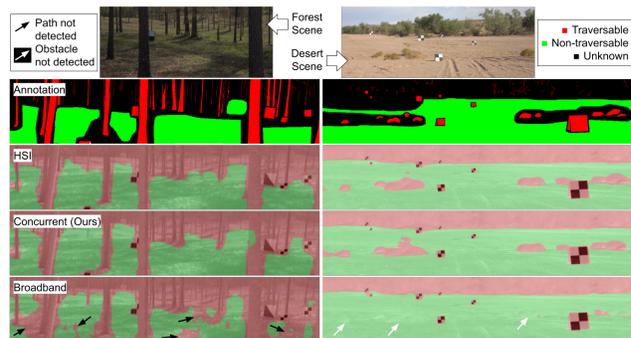


Figure 1. Using only 5 bands, our method is able to achieve traversability estimation similar to that achieved when using the full 256-band HSI. In contrast, broadband imagery often fails to detect traversable terrain and misses obstacles.

traversability estimation in off-road environments. However, as hyperspectral sensors can be costly, we focus on discovering the (potentially small) subset of bands that are most important for traversability estimation. Understanding how many and which bands have the largest impact on traversability estimation is necessary to inform the design of new, more cost-effective passive sensors with fewer bands.

We thus design a band selection module that can be used in conjunction with any segmentation network to discover an optimal subset of bands for traversability estimation. Our band selection module is implemented as a differentiable layer prepended to a segmentation network that learns to select a given number of bands to pass to the segmentation network. The band selection module is jointly trained with the segmentation network; when training is complete, the output is both a sensor design and a traversability estimation network that is co-optimized with that sensor design.

We validate our method on a dataset of diverse off-road environments, scenery, and weather conditions (see Fig. 1). The dataset, which will be released with this paper, was collected both during daytime and nighttime, using two different hyperspectral sensors (256-band and 250-band sensors). For both sensors, we demonstrate our ability to select 5 bands that can achieve similar traversability estimation performance to that achieved when using the full HSI.

2. Related Work

Our work lies at the intersection of off-road traversability, HSI segmentation, and machine learning for sensor design; we discuss related literature in this section.

Off-Road Traversability Estimation. Reina *et al.* [22] segment drivable terrain by classifying 3D point clouds as ground or non-ground according to their geometric properties, followed by radar-stereo integration for 3D obstacle detection. Leung *et al.* [15] leverage semantic plus geometric properties of the terrain to estimate its traversability. Hadsell *et al.* [9] determine long-range traversability of terrain by propagating labels backward in time; traversability labels from stereo-labeled image patches in the near range are used to classify the image patches in the far-range. Stavens *et al.* [26] propose a self-supervised approach to estimate terrain roughness from laser range data. Guan *et al.* [8] propose a geometric fusion approach to extract terrain features from RGB images and 3D point clouds, relying on semantic and geometric traversability estimations of terrain. Shaban *et al.* [24] estimate terrain traversability for autonomous off-road navigation by classifying the terrain into cost classes, using a recurrent neural network based architecture that predicts dense traversability maps from sparse LiDAR inputs. Zhu *et al.* [32] propose a reinforcement learning based method for terrain traversability analysis by exploiting behavior-based learning from demonstration methods by encoding vehicle kinematics into convolution kernels. Unlike the above, we focus on off-road traversability estimation using passive HSI sensors.

Hyperspectral Image Classification and Segmentation. HSI has become popular in a wide range of applications, from agriculture [4] to medicine [23]. Note that both “hyperspectral image classification” and “semantic segmentation from HSI” focus on classifying each pixel in an input hyperspectral image; the latter focuses on recognizing collections of pixels that form distinct categories, while the former does not necessarily assume that neighboring pixels typically belong to the same class. The two topics are closely related, so we discuss deep learning methods for both topics here. For a more detailed review, see [2].

There is a large body of work focused on hyperspectral image classification, but much of the work relies solely on spectral information and uses older, classical machine learning methods [13, 14, 16, 17, 29]. In recent years, development of deep learning methods for classification and semantic segmentation of HSI has begun [10, 20, 21, 30]. In [25] the authors propose ENL-FCN, a fully convolutional network with an efficient non-local module [12] for HSI classification. The convolutional network uses spectral-spatial information from a local region to classify each pixel in an input hyperspectral image, and the non-local module is designed to capture long-range contextual information for each pixel. In [31], a spectral-spatial dependent

global learning network (SSDGL-Net) is designed to combine global convolutional LSTM and global joint attention mechanism to classify pixels in HSI. In [23], the authors use a U-Net with pixel, superpixel, patch, and full image inputs for HSI segmentation. In [18], a deeply-supervised pseudo learning (DSPL) framework is used to capture multi-scale information. DSPL incorporates self-supervision by generating pseudo labels for unlabeled pixels and addresses class-imbalance with a pair-weighted loss. Finally, [11] introduces 3DSwinT, based on the self-attention transformer, for HSI classification using hierarchical contrastive learning.

Machine Learning for Sensor Configuration. Using machine learning to learn camera sensor configurations together with downstream vision tasks has recently been applied for a wide range of sensors and downstream tasks. In [5], the authors jointly learn a camera sensor’s color multiplexing pattern and a reconstruction network that attempts to reproduce the full color image via demosaicking. A “sensor layer” is used to select a single color channel per pixel by replacing the hard selection of color channels with a differentiable softmax with a “temperature” parameter. As training progresses and the temperature is increased, the soft selection increasingly resembles hard selection, and the multiplexing pattern is learned, together with the demosaicking network. Our method, inspired by [5], extends the idea of a sensor layer to address band selection in HSI while jointly learning a downstream segmentation network.

In [28], the authors address jointly optimizing a phase mask at the aperture plane of a camera and a downstream depth estimation network. The authors propose a differentiable optical layer whose learnable parameter is the height map of the phase mask. This layer takes an all-in-focus image together with a corresponding depth map and attempts to output a coded intensity image, which is used by the depth estimation network to estimate the depth map. [19] uses machine learning to learn an optical element that encodes high-dynamic-range (HDR) pixel information while simultaneously learning a decoding reconstruction network. The authors propose a “multiplexing” approach to HDR imaging that learns the diffractive optical element (DOE) that creates a point spread function (PSF) optimally suited for the downstream HDR image reconstruction network. Finally, [3] addresses single-shot hyperspectral-depth imaging. The authors jointly learn a DOE height map and a hyperspectral-depth reconstruction network. The DOE is used to simulate a PSF, which generates a simulated image, and the reconstruction network takes the simulated image and estimates the image spectrum and depth.

3. Concurrent Band Selection and Traversability Estimation from HSI

Given a dataset of HSI and associated pixel-level labels, our goal is to select the optimal, small subset of bands most

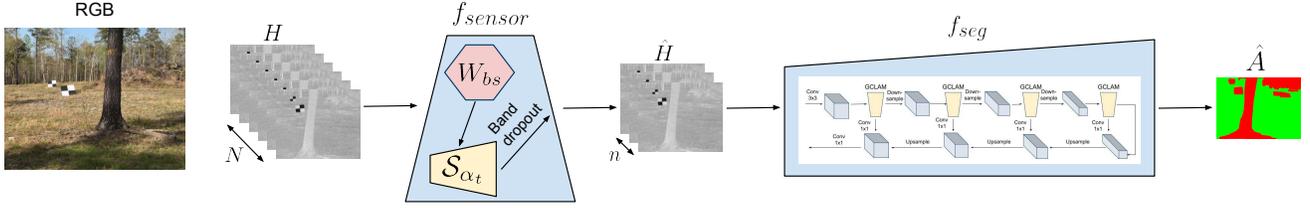


Figure 2. Concurrent Band Selection and Traversability Estimation architecture. Left: RGB image of the scene. Right: the band selection module—jointly trained with the segmentation network—chooses the optimal subset of bands for traversability segmentation in HSI.

important for traversability estimation, while jointly training a traversability estimation network, as shown in Fig. 2.

We pose the band selection problem as follows: assume we have a sensor with N bands, and we would like to select the $n \ll N$ bands that minimize a segmentation loss, where n is a known hyper-parameter. Assume we are given a collection of D N -band hyperspectral images $\{H_j(x, y) \in \mathbb{R}^N\}_{j=1}^D$ and associated pixel-level labels $\{A_j(x, y) \in \{1, 2, \dots, C\}\}_{j=1}^D$, where (x, y) indexes the spatial location in the image and there are C traversability labels (i.e. traversable, non-traversable, unknown).

We define a semantic segmentation network f_{seg} with parameters W_{seg} that estimates the traversability $\hat{A}(x, y)$ for each spatial location in a hyperspectral image,

$$\hat{A}(x, y) = f_{seg}(H(x, y); W_{seg}). \quad (1)$$

In addition to learning the parameters of the segmentation network, we would like to learn the parameters S of a function f_{sensor} that maps each input N -channel (spatial) pixel from H to an n -channel pixel $\hat{H}(x, y) \in \mathbb{R}^n$ via

$$\hat{H}(x, y) = f_{sensor}(H(x, y); S), \quad (2)$$

where $\hat{H}(x, y) \in \mathbb{R}^n$ mimics the pixel that would be acquired by a hypothetical n -band sensor. Importantly, we limit our hypothetical sensor design search space by assuming that the n bands are a subset of the N bands acquired by our real sensor. Expanding the search space to explore other band widths, band centers, etc. is left as future work.

We minimize the cross entropy segmentation loss \mathcal{L}_{seg} ,

$$\min_{W_{seg}, S} \sum_{j=1}^D \mathcal{L}_{seg}(f_{seg}(\hat{H}_j; W_{seg}), A_j) = \quad (3)$$

$$\min_{W_{seg}, S} \sum_{j=1}^D \mathcal{L}_{seg}(f_{seg}(f_{sensor}(H_j; S); W_{seg}), A_j),$$

jointly solving for the parameters of the segmentation network and the band selection module.

3.1. Band Selection Module

We implement the sensor function f_{sensor} in Eqn. 2 as

$$f_{sensor}(H(x, y); S) = S^T H(x, y) \quad (4)$$

$$\|s^i\| = 1 \quad \forall i,$$

where $S = [s^1, \dots, s^n] \in \{0, 1\}^{N \times n}$ is a binary variable and the norm of each column $\{s^i\}_{i=1}^n$ of S is constrained to

equal one. This results in f_{sensor} selecting exactly n of the N bands, so \hat{H} contains a subset of the original bands in H .

Learning the binary variable S is a challenging, combinatorial problem. Inspired by [5], we initially relax the binary constraint on S and iteratively converge on binary selections. Instead of directly learning S , we introduce a band selection parameter $W_{bs} = [w_{bs}^1, \dots, w_{bs}^n] \in \mathbb{R}^{N \times n}$ and set

$$s^i = \mathcal{S}_{\alpha_t}(w_{bs}^i), \quad (5)$$

where \mathcal{S}_{α_t} is the softmax operation with a scaling factor α_t . The hyper-parameter α_t increases quadratically with iteration t ; during the early stages of training we allow for linear combinations of the input bands, but by the end of training, each selected band corresponds to a unique input band. Specifically, we set $\alpha_t(t) = 1 + (\gamma t)^2$, where γ is a model hyperparameter. Notice that as α_t approaches infinity, S becomes a binary variable with $\|s^i\| = 1 \forall i$.

To encourage band diversity, so that each of the selected n bands corresponds to a different one of the N input bands, we use entropy regularization; we define the entropy regularization function

$$R(W_{bs}) = \frac{1}{N} \sum_{k=1}^N \mathcal{H}(w_{bs,k}), \quad (6)$$

where $\{w_{bs,k}\}_{k=1}^N$ denote the rows of W_{bs} and $\mathcal{H}(x) = -\sum \mathcal{S}(x) \log \mathcal{S}(x)$ is the entropy. We also add a band dropout layer, as in Fig. 2, after the band selection module to further encourage exploration of different solutions.

Finally, we jointly optimize the parameters of the segmentation network and the band selection module, replacing Eqn. 3 with

$$\min_{W_{seg}, W_{bs}} \sum_{j=1}^D \mathcal{L}_{seg}(f_{seg}(\hat{H}_j; W_{seg}), A_j) + \lambda R(W_{bs}) =$$

$$\min_{W_{seg}, W_{bs}} \sum_{j=1}^D \mathcal{L}_{seg}(f_{seg}(\mathcal{S}_{\alpha_t}(W_{bs})^T H_j; W_{seg}), A_j)$$

$$+ \lambda R(W_{bs}), \quad (7)$$

where λ is a regularization hyperparameter.

Fig. 3 shows the distributions of $\{s^i\}_{i=1}^n$ as the training progresses and α_t increases. Notice that as the training progresses, the distributions become more peaked. The regularization encourages the selection of unique bands, and

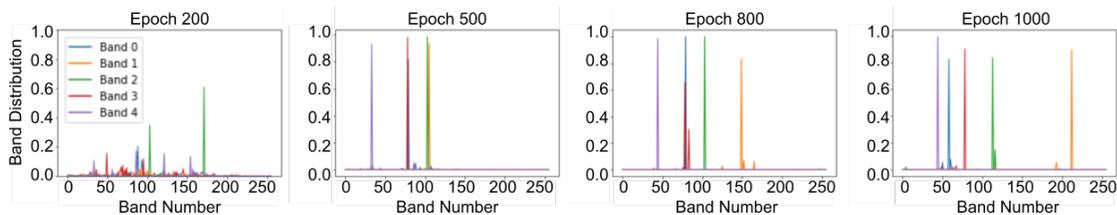


Figure 3. Band selection distributions for each of the 5 bands (shown in blue, orange, green, red, and purple) at various stages during training (after epochs 200, 500, 800, and 1000). As training progresses, the each distribution become peaked at a selected band.

by the end of the training process, $n = 5$ unique bands have been selected that span the $N = 256$ band spectrum.

4. Data

We evaluated our method on the Invisible Headlights (IH) Dataset (available at <https://www.kitware.com/ihdataset/>), which consists of simultaneously collected off-road LW HSI and visible imagery. The dataset was collected at multiple locations to capture wide variations in scenes and weather conditions, as described in Tab. 1. The IH Dataset was collected using a vehicle mounted with sensors: the broadband RGB sensor resolution is 3840×2160 , and the LW hyperspectral infrared resolution is 1500×260 or 1700×480 (see Tab. 1). The dataset was collected from different locations over different paths and steps. The hyperspectral sensors do not work on the move, so images were captured with the vehicle placed in set positions.

The collects range in scenery: Collect 1 consists of forested scenes and meadows, with trees, grass, water crossings, and gravel roads. Collect 2 has rocky desert lanes, sandy riverbeds, sparse and dense scrub brush and hardy plants, and shallow-banked dry riverbeds. Collect 3 consists of urban lanes, a closed hanger, a parking lot, train tracks, a runway with heavily and lightly traveled snow, snowy lanes through a forested road, and a fenced intersection. Collect 4 consists of sand, grass, asphalt, and pavement paths; fencing; a building; a pond; paths through a ditch; and paths through forest with small trees. For all collects, images were captured during various times of day and night.

Example images from different scene types are shown in Fig. 4. These images demonstrate some of the challenges in our dataset, such as a wooded scene where the vehicle faces the challenging task of navigating between the trees, a desert path that must be navigated during the nighttime, and scenes including water, snow, and various obstacles.

To train a supervised network to estimate traversability in HSI, we hand-annotated a subset of the imagery. The annotators had access to LW *broadband* (corresponding to LW *hyperspectral*) images, in addition to visual (RGB) images acquired from multiple viewpoints. Annotators labeled each pixel in the broadband images as traversable, non-traversable, or unknown; they were instructed to use best judgement as to whether a large vehicle would be able to traverse the terrain, and when unsure to label pixels as un-

known. Annotators used GIMP [1] to overlay on the imagery RGB annotation layers, which were transferred to PNG images where each pixel was labeled 0, 1, or 2, corresponding to unknown, traversable, and non-traversable. Examples of annotated images are shown in Fig. 5, and the annotated image statistics are provided in Tab. 1.

Because Collects 1 and 2 used a different LW HSI sensor than Collects 3 and 4, we split the IH Dataset into IH-256 (Collects 1 and 2, see Sec. 6.1) and IH-250 (Collects 3 and 4, see Sec. 6.2). This allows us to learn the optimal bands *per-sensor*, where each sensor captures bands with a different spectral resolution, with bands centered at different wavelengths. We further split the IH Dataset into training and testing splits, so that all imagery per path is in the same split. Details about the dataset splits are given in Tab. 2.

5. Experiments

We used the IH Dataset (Sec. 4) to train and evaluate our concurrent band selection and traversability estimation method and compared our method to several baselines. This section provides details about baselines and training details.

5.1. Traversability Estimation from HSI

We did preliminary experiments with three networks designed for HSI: ENL-FCN [25], SSDGL-Net [31], and a combination of the ENL segmentation head with the SSDGL-Net feature extractor. Based on these experiments, we selected SSDGL-Net as the best performing network on our data. SSDGL-Net, shown in Fig. 6, combines global convolutional LSTM with global joint attention to extract spectral and spatial correlations between pixels. We compute the cross entropy loss function with two output classes (traversable, non-traversable), ignoring unlabeled pixels.

5.2. Comparison with Other Methods

Full HSI. We expect that when utilizing *all* available image bands, the SSDGL network will be able to segment images with higher accuracy than when utilizing only a subset of the HSI bands. However, such a network requires significantly more model parameters and can be more difficult to train. We train the SSDGL network using the full hyperspectral image, without the band selection module, by optimizing $\min_{W_{seg}} \sum_{j=1}^D \mathcal{L}_{seg}(f_{seg}(H_j; W_{seg}), A_j)$.

Broadband. We simulate broadband imagery from our HSI dataset by summing all bands, similar to the simulation of a monochrome visible image by summing across RGB

Table 1. IH Dataset collect locations, image resolution, paths per collect, LW HSI per collect, and annotated HSI per collect.

Collect	Location	Image Resolution	Spectrum (microns)	# Paths	# Images	# Images Annotated	Season
1	Fort A.P. Hill, VA	1500 × 260 × 256	8.0 - 13.1	38	1,762	77	Spring
2	Sidewinder Range, YPG, AZ	1500 × 260 × 256	8.0 - 13.1	20	311	97	Summer
3	Loring Commerce Center, ME	1700 × 480 × 250	6.8 - 13.1	48	387	61	Winter
4	Avon Park Air Force Range, FL	1700 × 480 × 250	6.8 - 13.1	39	366	54	Spring
Total				145	2,826	289	N/A



Figure 4. Example IH Dataset scenes. Top, from left to right (IH-256): open field, wooded scene, desert (daytime), and desert (nighttime, infrared). Bottom, from left to right (IH-250): snowy scene, snowy scene with buildings, grassy scene with pond, grassy scene with trees. Targets in scenes were used for calibration purposes.

Table 2. Training and testing splits for the IH Dataset.

	IH-256		IH-250	
	Train	Test	Train	Test
# Paths	14	4	11	3
# Annotations	106	68	94	21

channels. While broadband imagery contains information from all HSI bands, it lacks spectral information that can lend insight into object materials, which could be useful for learning image semantics. To determine the impact of spectral information on semantic segmentation, we train the SSDGL network using simulated broadband imagery.

Random Bands. Our band selection experiments use only $n \ll N$ of the N available HSI bands. To evaluate the effectiveness of our band selection module, we train a SSDGL network using n randomly selected bands.

Evenly Spaced Bands. As with the Random Bands baseline, the Evenly Spaced Bands baseline evaluates the effectiveness of our band selection module by training a SSDGL network using only n bands. For this experiment, the n bands are selected by sampling evenly from the N bands.

Retraining Segmentation Network. This baseline assumes prior knowledge of the bands selected by our Concurrent Band Selection and Traversability Estimation (Concurrent) method. Assuming these bands are known, we train the SSDGL network using only these n bands as input. While we would expect the resulting performance to be similar to that obtained with our Concurrent method, the Retraining Segmentation Network (Retraining) baseline does *not* utilize the band selection module, as it assumes oracle knowledge of the “best” bands. This allows us to explore the impact of jointly training the band selection and

traversability estimation networks.

5.3. Training Details

Model hyperparameters. We use the SSDGL-Net with the hyperparameters as given in [31], but modify the network to work with our N -band, n -band, or 1-band imagery, depending on the experiment (the number of input bands differs for the various experiments, see Sec. 6). For band selection experiments, we selected $n = 5$ out of $N = 256$ or 250 bands, and we set the hyperparameter controlling the trade-off between the segmentation loss and band-diversity regularization loss to $\lambda = 1.0$. We set $\gamma = 20$ for the hyperparameter controlling the softmax scaling, and for the band dropout layer, we set the dropout probability to 0.8.

Data preparation and augmentation. During training, we use the annotated HSI dataset described in Sec. 4. We resize each image to have spatial dimension of 260×1600 (the image acquisition process results in slight image size variation), randomly rescale each image by a factor of 0.5 – 2.0, and randomly select crops of spatial dimension 128×256 . We use horizontal flipping augmentation and normalize images using the dataset mean and standard deviation. At test time, we only use image normalization.

Optimization and initialization. We train the band selection network in two stages, as in [5]. In the first stage, we jointly optimize parameters of the band selection module and segmentation network for 1000 epochs. We use the Adam optimizer with learning rates of 10^{-4} for the SSDGL-Net parameters and 10^{-3} for the band selection module, with weight decay of 10^{-5} . During the second stage of training, we freeze the band selection module parameters and finetune the SSDGL-Net parameters for 100

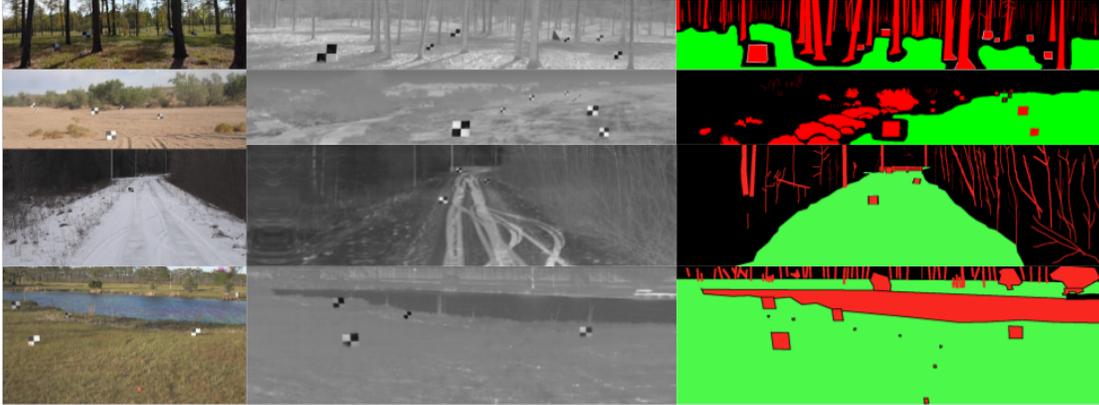


Figure 5. Example HSI annotations. Left: RGB visualization of four scenes (one per row). Middle: Broadband visualizations of associated hyperspectral images. Right: Associated traversability annotations, where red is non-traversable, green is traversable, and black is unknown / unlabeled. Note that the RGB imagery is not aligned with the HSI and is not used at all during training or evaluation.

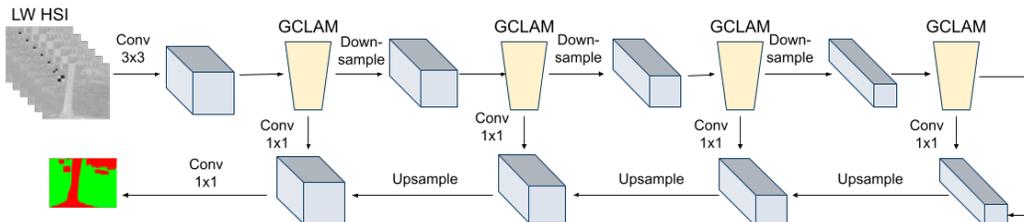


Figure 6. The SSDGL-Net architecture utilizes global convolutional LSTM integrated with a global joint attention mechanism (GCLAM); more details can be found in [31]. Skip connection are used to fuse spatial encoder features with semantic decoder features.

epochs. We use the Adam optimizer with an initial learning rate of 10^{-5} that decreases with a polynomial scheduler with a power of 0.9 until a minimum learning rate of 10^{-6} .

All baseline segmentation networks (without the band selection module) were trained for 1000 epochs using the Adam optimizer with initial learning rate of 10^{-4} and weight decay of 10^{-5} . The learning rate was decreased according to a polynomial learning rate scheduler with a power of 0.9 until a minimum learning rate of 10^{-6} .

All methods (including baselines) use a “pretrained” SSDGL backbone that was trained on a *subset* of the IH-256 Dataset. This is done for expediency, since we do not have access to any pretrained SSDGL backbones. Note that the segmentation head (a fully-connected network head) is trained from random initialization. All methods are trained with batch size 3 using a single NVIDIA RTX A6000 GPU.

6. Results

Recall that different sensors were used for the different data collects, so we split the IH Dataset into IH-256 and IH-250 (see Sec. 4). After training on the IH-256 and IH-250 Train splits, we use the Test splits (see Tab. 2) for evaluation of our method (Concurrent/Ours) compared to the baseline methods (Full HSI, Broadband, Random Bands, Evenly Spaced Bands, and Retraining) described in Sec. 5.2.

6.1. IH-256 Dataset

Our band selection module (Sec. 3) succeeds in selecting a subset of 5 bands that are unique, span across most

of the spectrum, and are able to achieve almost the same traversability estimation performance as when using the full HSI. Of the 256 bands (indexed 1, 2, . . . , 256), the selected bands are: 45, 59, 79, 114, and 213.

Fig. 8 shows qualitative results from example images, and a quantitative comparison of all methods is given in Tab. 3 and summarized in Fig. 7. Notice in Fig. 8 that the Broadband method sometimes labels traversable sand regions as non-traversable and also fails to label some brush as non-traversable. The Concurrent method attempts to distinguish sand regions from brush in the region encompassed by the black rectangle. Many of the methods are able to detect the branch, shown in the yellow square, and label it as non-traversable. Finally, notice that the results are similar for daytime (right) and nighttime (left).

Unsurprisingly, Full HSI has the best performance in terms of traversability estimation accuracy, with average pixel accuracy of 98.1% and average class accuracy of 97.22% (the discrepancy between pixel and class average is due to slightly worse performance on the non-traversable class). However, this model requires significantly more parameters than the other methods’ models, as seen in Fig. 7 (top right). The next-best method is our Concurrent method. Observe that the accuracies of traversable and non-traversable regions are only decreased by 0.39% and 0.03%, respectively, when using the five selected bands in place of the full 256-band HSI. Out of all methods that utilize only 5 out of the 256 bands, the methods that utilize the learned

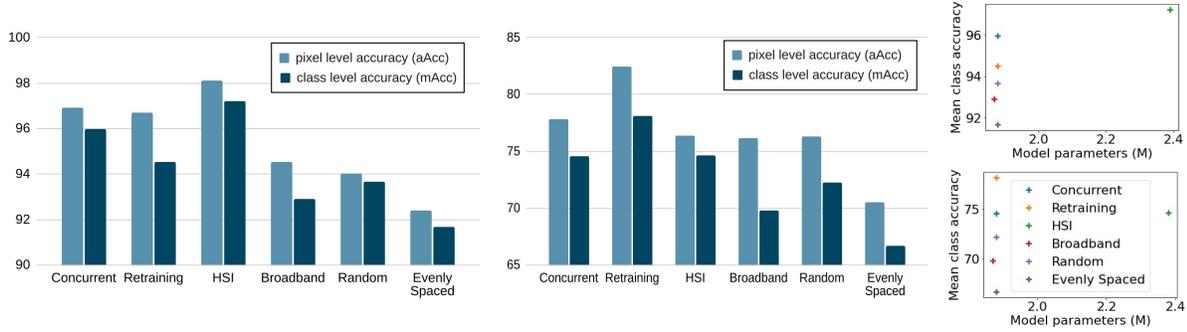


Figure 7. Comparison of all methods for the IH-256 (left) and IH-250 (middle) Datasets, and model size versus accuracy for the IH-256 (top right) and IH-250 (bottom right) Datasets.

Table 3. IH-256 results. Best method is bold; second best underlined. aAcc = average pixel-wise accuracy; mAcc = mean class accuracy.

Type	# Bands	Bands	Accuracy Traversable	Accuracy Non-Traversable	Summary	
					aAcc	mAcc
Concurrent	5	[44, 58, 78, 113, 212]	<u>96.60</u>	<u>95.32</u>	<u>96.91</u>	<u>95.96</u>
Retraining	5	[44, 58, 78, 113, 212]	96.02	93.02	96.70	94.52
HSI	255	[0, 1, ... 255]	98.27	96.18	98.12	97.22
Broadband	1	Sum[0, 1, ... 255]	93.20	92.63	94.53	92.92
Random	5	[37, 49, 106, 128, 246]	93.63	93.71	94.03	93.67
Evenly Spaced	5	[0, 64, 128, 191, 255]	88.66	94.70	92.40	91.68

5 bands (Concurrent and Retraining) perform best. Interestingly, jointly training the band selection and segmentation networks results in slightly better performance than training the segmentation network alone with the same bands.

The Random and Evenly Spaced methods also only utilize 5 bands; however these methods achieve mean class accuracies that are 2.29% and 4.28%, respectively, worse than our Concurrent method. This suggests that our band selection module is indeed selecting bands that are well-suited for traversability estimation. The naive method of sampling evenly across the spectrum is the worst-performing method, confirming the need for learning optimal bands for a given vision task. Finally, notice that the performance of the Broadband method is similar to that of the Random and Evenly Spaced methods, indicating that randomly or evenly sampling bands does not effectively utilize spectral information. Spectral information improved the traversable and non-traversable accuracies by at most 5.07% and 3.55%, respectively (the difference between the HSI and Broadband).

6.2. IH-250 Dataset

For the IH-250 Dataset, the band selection module again selects 5 unique bands that span most of the spectrum. Note that the sensors used to collect the IH-256 and IH-250 Datasets were different, and the band numbers from the IH-256 Dataset do *not* correspond to the same wavelengths in the IH-250 Dataset. Of the 250 bands, the selected bands were: 66, 85, 89, 116, and 139.

A quantitative comparison of all methods is given in Tab. 4 and summarized in Fig. 7. Notice the significant performance gap for *all* methods between the traversable and

non-traversable accuracies, and similarly between the average pixel and class accuracies. This indicates that the IH-250 Dataset, unlike the IH-256 Dataset, contains significantly more traversable than non-traversable pixels, and this dataset would benefit from class-balancing during training.

The best overall method is the Retraining method, which uses the “optimal” bands discovered by the Concurrent method. The mean class accuracy for the Retraining method is 78.1%, which is 3.5% better than that of the HSI method. The Concurrent method, which also utilizes the same 5 learned bands, performs slightly worse than the Retraining method, with a mean class accuracy of 74.6%, comparable with the performance of HSI method. While for the IH-250 Dataset the HSI method does not have the best overall performance, it has the second-best mean class accuracy of 74.6%. This could suggest that a longer training schedule is necessary to fully utilize the more complex spectral information contained in the 250 bands. Recall (Sec. 5.3), all methods were initialized with a backbone pretrained on a subset of the IH-256 Dataset; with additional training, the HSI method could possibly achieve better performance on IH-250, similar to its performance on the IH-256 Dataset.

The Random and Evenly Spaced methods, which only utilize 5 bands, achieve mean class accuracies that are 5.9% and 11.4% worse than the Retraining method, which uses 5 *learned* bands. This again suggests that our band selection module is choosing an optimal subset of the 250 available bands. Once again, the naive method of sampling evenly across the spectrum is the worst-performing method. The Broadband method suffers most from class imbalance, with the largest difference between traversable

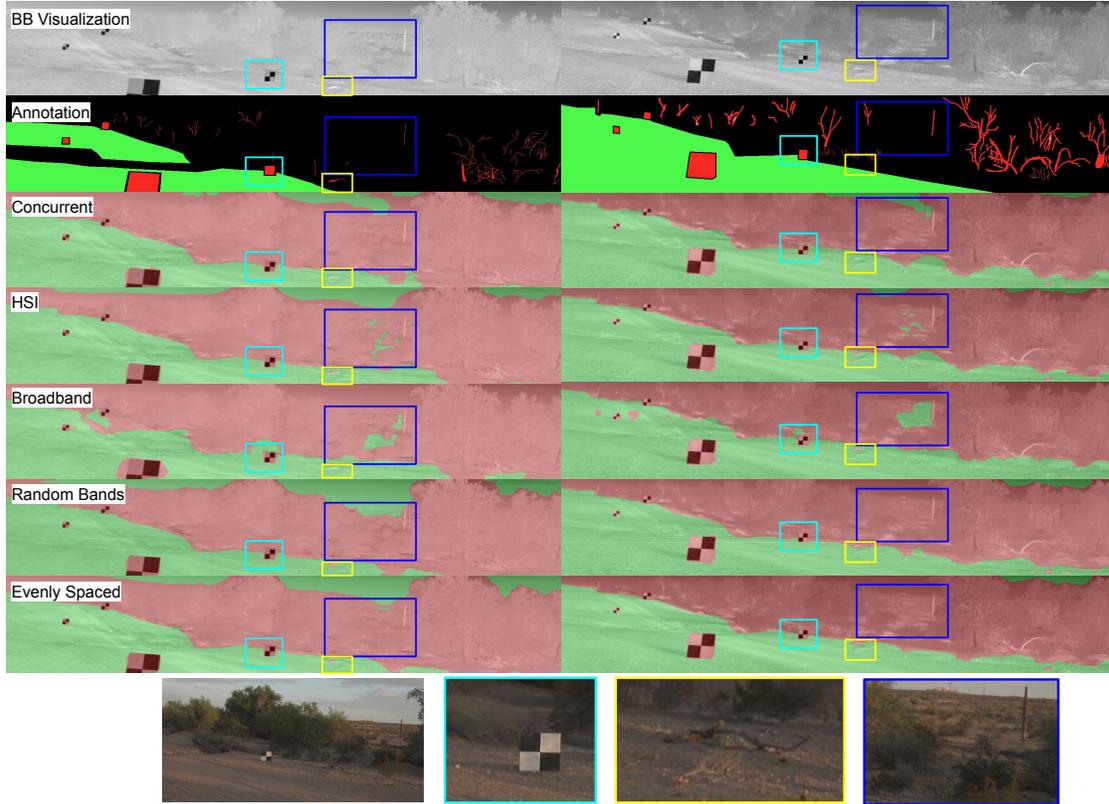


Figure 8. Example HSI traversability results of nighttime (left) and daytime (right) paths from the same desert scene. Bottom: Daytime RGB image of the scene. Red = non-traversable, green = traversable.

Table 4. IH-250 results. Best method is bold; second best underlined. aAcc = average pixel-wise accuracy; mAcc = mean class accuracy.

Type	# Bands	Bands	Accuracy Traversable	Accuracy Non-Traversable	Summary	
					aAcc	mAcc
Concurrent	5	[66, 85, 89, 116, 139]	81.46	67.69	<u>77.81</u>	74.57
Retraining	5	[66, 85, 89, 116, 139]	85.45	<u>70.81</u>	82.44	78.13
HSI	255	[0, 1, ... 249]	77.65	71.61	76.36	<u>74.62</u>
Broadband	1	Sum[$\{0, 1, \dots 249\}$]	<u>82.27</u>	57.37	76.18	69.82
Random	5	[7, 117, 140, 191, 225]	79.96	64.49	76.27	72.22
Evenly Spaced	5	[0, 62, 124, 178, 249]	71.39	62.05	70.54	66.72

and non-traversable accuracies. The mean class accuracy for Broadband is again similar to that of the Random and Evenly Spaced methods. Finally, we note that the overall increase in mean class accuracy when using HSI compared to Broadband is 4.8%, but when retraining with the 5 selected bands, the increase in accuracy is 8.3%. This suggests that training SSGDL-Net to utilize the full spectrum is more challenging than training it with a small subset of bands, despite its being designed for HSI classification.

7. Conclusion

We have demonstrated the feasibility of using passive LW hyperspectral sensors for autonomous navigation in diverse off-road environments. We have shown that our band selection method selects a small subset of optimal bands from two different hyperspectral sensors, suggestive for

the design of new, more affordable passive sensors for autonomous navigation. Also, we have shown that while we can select just five bands without dramatically impacting performance, using a single broadband channel does not provide sufficient signal for autonomous navigation. Future work includes addressing class imbalance in our dataset and optimization over the number of bands selected. Finally, we note that our method can be applied not only to other sensors but also for other downstream vision tasks.

Acknowledgement This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References

- [1] Gimp gnu image manipulation program. 4
- [2] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. Deep learning for classification of hyperspectral data: A comparative review. *IEEE geoscience and remote sensing magazine*, 7(2):159–173, 2019. 2
- [3] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2651–2660, 2021. 2
- [4] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, Sep 2015. 2
- [5] Ayan Chakrabarti. Learning sensor multiplexing design through back-propagation. *Advances in Neural Information Processing Systems*, 29, 2016. 2, 3, 5
- [6] Christopher Goodin, Lalitha Dabiru, Christopher Hudson, George Mason, Daniel Carruth, and Matt Doude. Fast terrain traversability estimation with terrestrial lidar in off-road autonomous navigation. In *Unmanned Systems Technology XXIII*, volume 11758, pages 189–199. SPIE, 2021. 1
- [7] Christopher Goodin, Matthew Doude, Christopher R Hudson, and Daniel W Carruth. Enabling off-road autonomous navigation-simulation of lidar in dense vegetation. *Electronics*, 7(9):154, 2018. 1
- [8] Tianrui Guan, Zhenpeng He, Dinesh Manocha, and Liangjun Zhang. TTM: terrain traversability mapping for autonomous excavator navigation in unstructured environments. *CoRR*, abs/2109.06250, 2021. 2
- [9] Raia Hadsell, Pierre Sermanet, Ayse Naz Erkan, Jan Ben, Jefferson Han, Beat Flepp, Urs Muller, and Yann LeCun. Online learning for offroad robots: Using spatial label propagation to learn long-range traversability. In Wolfram Burgard, Oliver Brock, and Cyrill Stachniss, editors, *Robotics, Robotics: Science and Systems*, pages 17–23. MIT Press Journals, 2008. 3rd International Conference on Robotics Science and Systems, RSS 2007 ; Conference date: 27-06-2007 Through 30-06-2007. 2
- [10] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015:1–12, 2015. 2
- [11] Xin Huang, Mengjie Dong, Jiayi Li, and Xian Guo. A 3-d-swin transformer-based hierarchical contrastive learning method for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. 2
- [12] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. Ccnet: Criss-cross attention for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 603–612, 2019. 2
- [13] Sen Jia, Linlin Shen, Jiasong Zhu, and Qingquan Li. A 3-d gabor phase-based coding and matching framework for hyperspectral imagery classification. *IEEE transactions on cybernetics*, 48(4):1176–1188, 2017. 2
- [14] Xudong Kang, Shutao Li, and Jon Atli Benediktsson. Spectral–spatial hyperspectral image classification with edge-preserving filtering. *IEEE transactions on geoscience and remote sensing*, 52(5):2666–2677, 2013. 2
- [15] Tiga Ho Yin Leung, Dmitry Ignatyev, and Argyrios Zolotas. Hybrid terrain traversability analysis in off-road environments. In *2022 8th International Conference on Automation, Robotics and Applications (ICARA)*, pages 50–56, 2022. 2
- [16] Jun Li, José M Bioucas-Dias, and Antonio Plaza. Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3):809–823, 2011. 2
- [17] Ting Lu, Shutao Li, Leyuan Fang, Xiuping Jia, and Jón Atli Benediktsson. From subpixel to superpixel: A novel fusion framework for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4398–4411, 2017. 2
- [18] Weiran Luo, Chengcai Zhang, Ying Li, Feng Yang, Dongying Zhang, and Zhiming Hong. Deeply-supervised pseudo learning with small class-imbalanced samples for hyperspectral image classification. *International Journal of Applied Earth Observation and Geoinformation*, 112:102949, 2022. 2
- [19] Christopher A Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1375–1385, 2020. 2
- [20] Lichao Mou, Pedram Ghamisi, and Xiao Xiang Zhu. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3639–3655, 2017. 2
- [21] Mercedes E Paoletti, Juan Mario Haut, Ruben Fernandez-Beltran, Javier Plaza, Antonio J Plaza, and Filiberto Pla. Deep pyramidal residual networks for spectral–spatial hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):740–754, 2018. 2
- [22] Giulio Reina, Annalisa Milella, and Raphaël Rouveure. Traversability analysis for off-road vehicles using stereo and radar data. In *IEEE International Conference on Industrial Technology, ICIT 2015, Seville, Spain, March 17-19, 2015*, pages 540–546. IEEE, 2015. 2
- [23] Silvia Seidlitz, Jan Sellner, Jan Odenthal, Berkin Özdemir, Alexander Studier-Fischer, Samuel Knödler, Leonardo Ayala, Tim J Adler, Hannes G Kenngott, Minu Tizabi, et al. Robust deep learning-based semantic organ segmentation in hyperspectral images. *Medical Image Analysis*, 80:102488, 2022. 2
- [24] Amirreza Shaban, Xiangyun Meng, JoonHo Lee, Byron Boots, and Dieter Fox. Semantic terrain classification for off-road autonomous driving. In Aleksandra Faust, David Hsu, and Gerhard Neumann, editors, *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pages 619–629. PMLR, 08–11 Nov 2022. 2
- [25] Yu Shen, Sijie Zhu, Chen Chen, Qian Du, Liang Xiao, Jianyu Chen, and Delu Pan. Efficient deep learning of nonlocal fea-

- tures for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):6029–6043, 2020. [2](#), [4](#)
- [26] David Stavens and Sebastian Thrun. A self-supervised terrain roughness estimator for off-road autonomous driving. *CoRR*, abs/1206.6872, 2012. [2](#)
- [27] Maggie Wigness, Sungmin Eum, John G Rogers, David Han, and Heesung Kwon. A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5000–5007. IEEE, 2019. [1](#)
- [28] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. Phasecam3d—learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2019. [2](#)
- [29] Haoyang Yu, Lianru Gao, Wenzhi Liao, Bing Zhang, Lina Zhuang, Meiping Song, and Jocelyn Chanussot. Global spatial and local spectral similarity-based manifold learning group sparse representation for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5):3043–3056, 2019. [2](#)
- [30] Zhuo Zheng, Yanfei Zhong, Ailong Ma, and Liangpei Zhang. Fpga: Fast patch-free global learning framework for fully end-to-end hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(8):5612–5626, 2020. [2](#)
- [31] Qiqi Zhu, Weihuan Deng, Zhuo Zheng, Yanfei Zhong, Qingfeng Guan, Weihua Lin, Liangpei Zhang, and Deren Li. A spectral-spatial-dependent global learning framework for insufficient and imbalanced hyperspectral image classification. *IEEE Transactions on Cybernetics*, 2021. [2](#), [4](#), [5](#), [6](#)
- [32] Zeyu Zhu, Nan Li, Ruoyu Sun, Huijing Zhao, and Donghao Xu. Off-road autonomous vehicles traversability analysis and trajectory planning based on deep inverse reinforcement learning. *CoRR*, abs/1909.06953, 2019. [2](#)