

Hyperspectral CNN Classification with Limited Training Samples

Lloyd Windrim

l.windrim@acfr.usyd.edu.au

Rishi Ramakrishnan

r.ramakrishnan@acfr.usyd.edu.au

Arman Melkumyan

a.melkumyan@acfr.usyd.edu.au

Richard J. Murphy

richard.murphy@sydney.edu.au

Australian Centre for Field Robotics

University of Sydney

Sydney, Australia

Abstract

Hyperspectral imaging sensors are becoming increasingly popular in robotics applications such as agriculture and mining, and allow per-pixel thematic classification of materials in a scene based on their unique spectral signatures. Recently, convolutional neural networks have shown remarkable performance for classification tasks, but require substantial amounts of labelled training data. This data must sufficiently cover the variability expected to be encountered in the environment. For hyperspectral data, one of the main variations encountered outdoors is due to incident illumination, which can change in spectral shape and intensity depending on the scene geometry. For example, regions occluded from the sun have a lower intensity and their incident irradiance skewed towards shorter wavelengths.

In this work, a data augmentation strategy based on relighting is used during training of a hyperspectral convolutional neural network. It allows training to occur in the outdoor environment given only a small labelled region, which does not need to sufficiently represent the geometric variability of the entire scene. This is important for applications where obtaining large amounts of training data is labourious, hazardous or difficult, such as labelling pixels within shadows. Experiments to evaluate the classification system are carried out on three outdoor datasets captured from a field-based platform.

1 Introduction

Classification algorithms which use hyperspectral data form a critical component of outdoor robotic systems as the spectral signature of an object is its most defining characteristic. While the majority of robotic platforms contain consumer grade cameras that have broadband spectral response curves, field based hyperspectral imagers are becoming increasingly common in applications such as mineral mapping of mine faces [19], weed detection in agriculture [23], urban imaging [18] and skin detection in search and rescue [2]. The chemical and structural composition of a material gives rise to unique spectral signatures, allowing per-pixel thematic classification maps of the environment to be generated through the use of

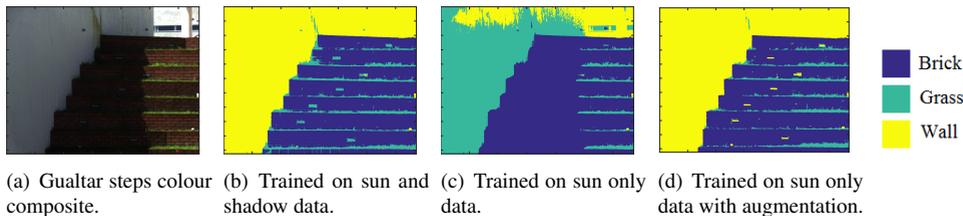


Figure 1: Improvement in CNN per-pixel classification of a hyperspectral image with a shadow when using spectral relighting augmentation. Training labels were extracted from either the entire image or only sunny areas. Results shown for the Gualtar steps dataset.

supervised classification algorithms. Classifiers require labelled data to train on, but acquiring large scale ground truth labels is expensive in terms of computation and labelling effort. In such scenarios, it is desirable to label small regions within the image and infer class labels on the remaining pixels with the classifier.

Incorporating illumination variability into the training of supervised classifiers is an ongoing research question, with two main approaches being utilised. The first is by using large amounts of training data to sufficiently represent the variability in the scene. However, this is difficult in complex scenes where the incident illumination varies on a per-pixel basis. The second method is to use pre-processing to convert the data to a form that is less dependant on illumination. This is typically done by normalising the raw digital values against a material of known reflectance such as a calibration board (flat-field correction). However, this is only correct for the region in which the calibration board was placed [17].

Convolutional Neural Networks (CNNs) have been utilised to achieve state-of-the-art performance for tasks such as classification [2, 9, 21] and robotic applications [16, 20]. Recently, they have also been applied to hyperspectral data captured from airborne and satellite based platforms [25].

In this work, CNN architectures are harnessed to generate an outdoor hyperspectral classification system from limited amounts of labelled training data. Illumination variability is incorporated into the classifier through the use of a data augmentation strategy that uses relighting. The advantages of the proposed approach are that it does not require multiple sensor modalities or computational atmospheric models. It allows training data to be sampled from within sunlit regions which are commonly easier to label, while classification can be performed on both sunlit and shadowed regions (Figure 1). The contributions of this paper are:

- an automatic, image based method for determining the incident illumination ratio,
- the incorporation of relighting into data augmentation,
- evaluation on outdoor hyperspectral datasets captured from a field-based platform.

Section 2 discusses related work and Section 3 describes the proposed method. Experiments and results are presented in Section 4, and conclusions are drawn in Section 5.

2 Previous Work

2.1 Illumination Invariance in Hyperspectral Classification

In the remote-sensing community, a well-known classifier is the Spectral Angle Mapper (SAM), which computes the inverse cosine of the normalised dot product between a spectra of interest and reference spectra [24]. SAM achieves some degree of illumination invariance as the classifier is robust to multiplicative scalings of the spectra that are constant across wavelength. This does not cater for the effect of shadows which are a wavelength dependant multiplication of the spectra. An alternative approach is to transform the spectra into an illumination invariant form prior to classification. From computer vision, the method of [5] derives a transformation for RGB images to a 2D log-chromaticity space in which an axis exists where changes in illumination due to intensity and shadow are suppressed. This approach is extended to hyperspectral images in [4]. The problem with this approach is in the assumptions that are made in order to derive the transformation, particularly the assumption of Planckian illumination where the spectral power distribution of the incident light is modelled by Wein's approximation to Plank's law. Atmospheric absorption features are not accounted for by the approximation even though they have a large impact on the incident light spectra. Finally, multi-modal approaches [7, 18] which use additional sensors such as Light Detection and Ranging (LiDAR) and Global Positioning Systems (GPS) can form geometry-based illumination models of the scene and compensate for the variations in lighting, however, the additional sensors required in these approaches are not always available.

2.2 CNNs for Hyperspectral Classification

In recent years, CNNs have been utilised for pixel-wise classification of hyperspectral images. However, many of the proposed CNNs convolve in the spatial domain and not the spectral domain. The CNN in [12] convolves a small window over spatial patches extracted from a dimensionality reduced hyperspectral cube. A similar approach is taken by [25], but the spatial features are combined with spectral features learnt using a method based on Local Discriminant Embedding. In [10] spatial patches are extracted from a non-reduced hyperspectral cube, but there are still no convolutions occurring over the spectral channel. In scenarios such as the ones presented in this paper where the amount of labelled training is heavily constrained, it is very difficult to convolve in the spatial domain where the data is limited, and so attention must be focussed on convolving in the information-rich spectral domain. A very simple CNN was proposed by [8] which learnt features by convolving over the spectral channel. This approach was shown to perform favourably against other types of neural networks as well as SVMs. In all of these works all of the scenarios assume that the training examples are sufficient for capturing the variability in the data, which is not always possible in a robotics or autonomous application. Also, the datasets used for evaluation were almost exclusively captured from satellite or airborne platforms.

3 A Robust Hyperspectral CNN

In this section, a summary of how to train the proposed network is given, followed by the data augmentation strategy that makes the classifier robust to illumination variability. Critical

to this, is the calculation of the ratio between the two primary outdoor illumination sources (terrestrial sunlight and diffuse skylight), for which a novel, image based method is proposed.

The proposed hyperspectral CNN training strategy used to increase the robustness of the classifier to illumination variability consists of multiple convolutional and non-linear activation layers, followed by fully connected layers and the output softmax classifier. Each pixel spectra is a data point and the network is trained by convolving filters over the entire spectrum. There are no pooling layers so that the location of the spectral features along the spectrum is preserved. The network is trained using stochastic gradient descent optimisation. To achieve robustness to illumination variability, each batch of training data is augmented with relighting and passed into the CNN. Convolution is only performed in the spectral domain. This is because texture can be an unreliable feature, especially in the scenarios dealt with in this work where the number of labels is limited.

3.1 Spectral Relighting Augmentation

Training supervised classification algorithms using small labelled regions fails to account for illumination variability induced by the complex geometry of a scene. This paper proposes the use of relighting [18] as a data augmentation strategy, in order to encompass the illumination variations typically found in the outdoor environment. This allows training data to be obtained from regions in the image that are either easy to access or easy to label, and inference can then be performed on the remaining data. The following relighting derivations focus on obtaining labels from sunlit regions, and classifying on shadowed data, though the approach is easily transferable to the reverse scenario.

The outdoor illumination model [18] consists of a parallel, terrestrial sunlight source $\mathbf{E}_{sun}\tau$, and a hemispherical diffuse skylight source \mathbf{E}_{sky} . Assuming all materials in the scene diffusely reflect light and that indirect illumination is negligible, the radiance L of a region i as captured by a camera can be approximated as:

$$L_i(\lambda) = \frac{\rho_i}{\pi} [V_i E_{sun}(\lambda)\tau(\lambda)\cos\theta_i + \Gamma_i E_{sky}(\lambda)], \quad (1)$$

where ρ_i is the albedo of the material, V_i is a binary variable indicating whether there is line-of-sight visibility between the region and the sun position, θ_i is the angle between the surface normal N_i and the vector towards the sun, and Γ_i is the sky (or view) factor ranging from 0 to 1 indicating the portion of the sky dome that is visible.

Relighting is the process of simulating the spectral appearance of a region under different illumination and geometrical conditions that are not encompassed by the training set. For a single sunlit datapoint at region i , the augmented version L_j is calculated by multiplying the training spectra L_i by a wavelength dependent scaling factor:

$$L_j(\lambda) = L_i(\lambda) \frac{V_j \frac{E_{sun}(\lambda)\tau(\lambda)}{E_{sky}(\lambda)} \cos\theta_j + \Gamma_j}{\frac{E_{sun}(\lambda)\tau(\lambda)}{E_{sky}(\lambda)} \cos\theta_i + \Gamma_i}, \quad (2)$$

where θ_i and Γ_i are the geometric parameters describing the sun angle and sky factor of the original training datapoint, while V_j , θ_j and Γ_j are the parameters of the augmented datapoint. When V_j is 0, relighting has the effect of simulating a shadowed region, while setting it to 1 simulates the same datapoint with a different orientation.

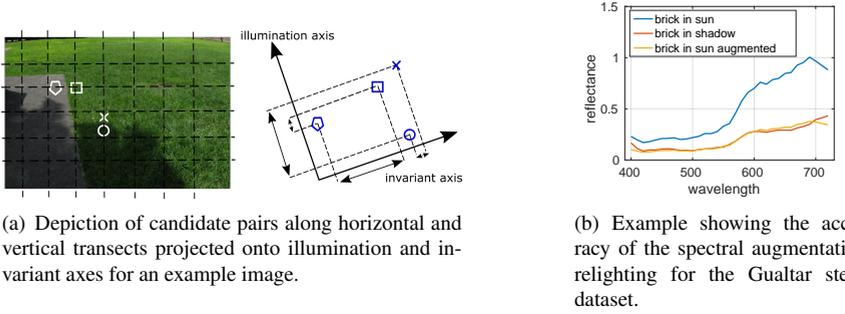


Figure 2: Extraction of the terrestrial sunlight-diffuse skylight ratio and its application in augmenting spectra to appear in shadow.

In order to relight the data during training, several illumination and geometric parameters are required. The first is the terrestrial sunlight-diffuse skylight ratio $\frac{E_{sun}(\lambda)\tau(\lambda)}{E_{sky}(\lambda)}$, which describes the spectral and intensity relationship between the primary illumination sources in the outdoor environment. The geometric parameters such as the sun angle and sky factors are typically known when utilising multi-modal systems, where geo-registered point cloud data can be used explicitly to estimate these values [18]. However, for image based methods these parameters remain unknown, therefore a sampling procedure is used during training as described in Section 3.2. During each batch of gradient descent optimisation of the network, the geometric parameters $\theta_A, \Gamma_A, V_j, \theta_j, \Gamma_j, \theta_i$ and Γ_i are sampled as shown in Algorithm 1.

Algorithm 1: Augmenting a batch of spectra for training the CNN. \mathcal{U} and \mathcal{B} represent uniform and Bernoulli distributions respectively.

Input : batch of spectra **data**, number of irradiance ratio estimates M
Output: augmented batch of spectra **dataAug**

```

1 dataAug  $\leftarrow$  data
2 for  $k = 1$  to  $M$  do
3   sample  $\theta_A \sim \mathcal{U}[0, \frac{\pi}{2})$ ,  $\Gamma_A \sim \mathcal{U}[0, 1]$ 
4   irradianceRatio  $\leftarrow$  eq.(3)( $\theta_A, \Gamma_A$ )
5   for  $l = 1$  to size of batch do
6      $V_j \sim \mathcal{B}(1, \frac{1}{2})$ ,  $\theta_i \sim \mathcal{U}[0, \frac{\pi}{2})$ ,  $\theta_j \sim \mathcal{U}[0, \frac{\pi}{2}]$ ,  $\Gamma_i \sim \mathcal{U}[0, 1]$ ,  $\Gamma_j \sim \mathcal{U}[0, 1]$ 
7     relictSpectra  $\leftarrow$  eq.(2)(data;  $V_j, \theta_i, \theta_j, \Gamma_i, \Gamma_j, \mathbf{irradianceRatio}$ )
8     dataAug  $\leftarrow$  dataAug  $\cup$  relictSpectra
9   end
10 end

```

3.2 Image Based Estimation of the Terrestrial Sunlight-Diffuse Skylight Ratio

The terrestrial sunlight-diffuse skylight ratio is integral to the relighting process. Manual methods involve the user selecting two adjacent points obtained from the same material [18], or the use of computational atmospheric models, but these require the knowledge of

parameters such as turbidity, gas concentration and water vapour [10]. Automatic methods also exist [11], but these require extra sensor modalities (e.g. LiDAR). In this work, a novel image based method for estimating the terrestrial sunlight-diffuse skylight ratio is proposed. It consists of a three stage process of candidate generation, refinement and smoothing.

If two pixels (A, A') of the same material are selected from a sunlit and shadowed region respectively, both of which have the same orientation, the terrestrial sunlight-diffuse skylight ratio can be approximated by equating ρ_A and $\rho_{A'}$ from (1):

$$\frac{E_{sun}(\lambda)\tau(\lambda)}{E_{sky}(\lambda)} = \frac{\Gamma_A}{\cos \theta_A} \left[\frac{L_A(\lambda)}{L_{A'}(\lambda)} - 1 \right] \propto \frac{L_A(\lambda)}{L_{A'}(\lambda)} - 1. \quad (3)$$

Since the scene geometry is unknown due to the use of only image data, selection of pairs of points in sunlit and shadowed regions from the same material can be used to obtain candidate terrestrial sunlight-diffuse skylight ratios that will be a scalar multiple of the underlying ratio.

To automatically select pairs of points in and out of the shadowed regions, a pseudo RGB image is formed from three bands. Hypothetically these can be any bands, however they should be chosen to maximise the discriminability of the classes. If in the visible domain, these bands can be 450nm, 550nm and 600nm (peak wavelengths of an RGB camera), and if in the Short-Wave Infrared (SWIR) domain the bands can be 1060nm, 1250nm and 1630nm (the middle of sections of the spectrum outside the destructive water bands). Next, the three channel image is converted to 2D log-chromaticity space where the illumination invariant direction is found through entropy minimization [9, 6]. The 1D projection of the image onto this axis is invariant to changes in the illumination. The orthogonal axis is found (deemed the illumination axis) which captures large changes in illumination resulting from either shadow or spectrally discrete class boundaries.

Candidate pairs of adjacent points taken from horizontal and vertical transects of the pseudo RGB image are projected onto the invariant and illumination axis:

$$I_{inv_i} = e^{X_i w}, I_{ill_i} = e^{X_i w^\perp}, \quad (4)$$

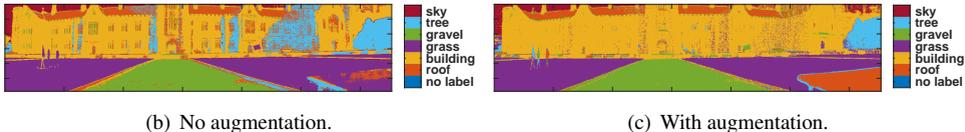
where X_i is the log-chromaticity of point i , the vector w is the direction of the invariant axis and I_i is the exponential of the points location on either the invariant or illumination axis. If there is a large difference between a pair along the illumination axis (corresponding to either a shadow or material class boundary) but a small difference between the pair along the invariant axis (ruling out the class boundary), then the pair is considered to be valid (Figure 2(a)) and the ratio between the spectra is calculated. For a candidate pair of points, the validity of them constituting a sun-shadow pair can be determined using:

$$\frac{|I_{inv_1} - I_{inv_2}|}{I_{inv_2}} < \mu, \frac{|I_{ill_1} - I_{ill_2}|}{\min(I_{ill_1}, I_{ill_2})} > \xi, \quad (5)$$

where reasonable values for μ and ξ were determined empirically as 0.3 and 1.2 respectively. The average ratio is subsequently taken over all valid candidate pairs before smoothing with an Savitzky-Golay filter [12] and used as an estimate for $\frac{L_A(\lambda)}{L_{A'}(\lambda)}$ in (3).



(a) False colour image of Great Hall building with example of limited sunlit regions where labels are extracted.



(b) No augmentation.

(c) With augmentation.

Figure 3: CNN classification of the Great Hall using 1000 training examples per class were used, drawn from the limited regions (red squares).

4 Experimental Method and Results

4.1 Datasets

Three outdoor field-based datasets captured from hyperspectral sensors are used to evaluate the proposed relighting augmentation. The first dataset is a SWIR scan of the Great Hall at the University of Sydney (Figure 3(a)) [18]. The image consists of 293×1306 pixels with 152 spectral channels ranging from 1009 – 2482nm. The scene consists of an urban environment, with material classes including roof, building, grass, path, tree and sky. Large shadows exist, predominantly over the building and roof regions. The second dataset is a Visible and Near-Infrared (VNIR) scan of a mine face [14], exemplifying a natural, unstructured environment. It contains two types of ore; Martite and Shale and a geologist has identified a rough boundary separating the two geological regions (Figures 4(a) and 4(b)). Two scans are captured from the same position at different times of the day, one at 11:30 and one at 13:30, with shadows emerging in the latter image. Each image is 289×1443 pixels, with 220 spectral channels corresponding to wavelengths in the range of 401-970nm. The third dataset consists of a Visible (VIS) range image of a semi urban scene in Gualtar [15]. The image is 1024×1344 pixels, with 33 channels in the range 400-720nm and classes of brick, grass and wall (Figure 1(a)). Half of the image is in shadow.

4.2 CNN architecture

The overall CNN architecture is the same for all datasets. Empirically, it was determined that the best architecture contained two convolutional and two fully connected layers, with the first convolutional learning 30 filters and always having a filter size of $1 \times 1 \times 30 \times 1$ (each input spectra was of shape $1 \times 1 \times D \times 1$ where D is the number of spectral channels). Since the Gualtar steps data only has 33 channels, the filter size used is $1 \times 1 \times 10 \times 1$. The second convolutional layer has ten filters each with a size of $1 \times 1 \times 10 \times 30$. Fully connected layers consist of 20 units each. The networks are optimised using 200 epochs of stochastic gradient descent with a learning rate of 10^{-5} , momentum of 0.9 and batch size of 50.

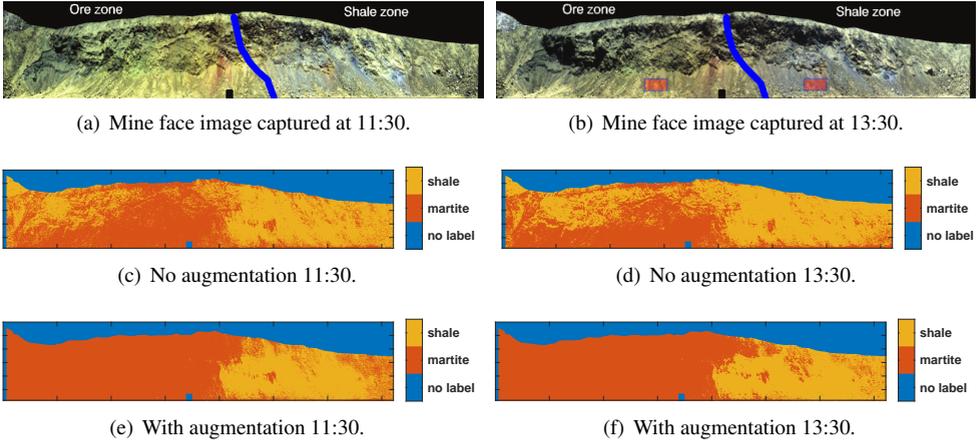


Figure 4: CNN classification of the mine face using 100 training examples per class drawn from the limited regions (red squares). The same network was used to classify the image at both times of the day. The blue line is a rough class boundary identified by geologists.

4.3 Results

To assess the performance of the CNN with the spectral relighting augmentation for each of the experiments, the F1 score is used. To evaluate the classifiers, test set sizes of roughly 225000, 90000 and 230000 pixels spectra are chosen for the Great Hall, mine face and Gualtar steps datasets respectively. The number of training examples is varied logarithmically between 100 and 1000 examples per class (5 increments: 100, 178, 316, 562, 1000). Ten candidate terrestrial sunlight-diffuse skylight ratios are generated for the augmentation such that each pixel spectra is relit ten times, with roughly half of those relightings being to shadow and half remaining sunlit but with different orientations. This expands the training dataset to 11 times its original size. The validation dataset consists of 50 examples per class. The mean and standard deviation for five randomly initialised networks is recorded.

Separate CNNs are trained on two sets of labelled training data. One set (referred to as ‘comprehensive’) is collected from both sunlit and shaded regions of the image, and also has large spatial coverage such that it best represents the variation in scene geometry and incident illumination. The second labelled training dataset (referred to as ‘limited’) is only collected in sunlit regions and the spatial coverage is small - limited to a patch (red squares in Figures 3(a) and 4(b)), such that there is a very poor representation of the scene geometry and incident illumination. Networks are trained on the limited dataset using the proposed data augmentation. Their performance is compared to networks that are trained on the limited and comprehensive datasets without any augmentation. The results of the networks trained on the comprehensive training data can be seen as roughly an upper bound for the results of training on the limited datasets. The performance of the CNN trained on the limited data, both augmented and not augmented, is compared against other classification approaches including SAM [24], an SVM [3], and a CNN trained on spectra projected into an illumination invariant space using log-chromaticity [4]. For the mine face dataset, networks are trained on the 13:30 image only.

The results (Figures 1, 3, 4 and 5) from all three datasets showed that there was a clear advantage to augmenting the labelled data used to train the CNN. The CNN achieved better

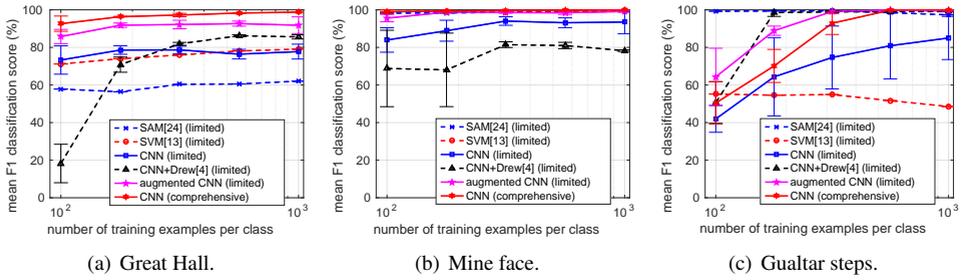


Figure 5: Spectral Relighting Augmentation Results: Results showing the benefit of training a CNN with the proposed data augmentation approach. The performance of the augmented CNN is compared with other approaches. 'Limited' indicates that a classifier was trained on localised, well-lit regions of the image, and 'comprehensive' indicates that a classifier was trained on the whole image (both well-lit and shadowed regions, see Figures 3(a) and 4(b)). The number of training samples is varied logarithmically between 100 and 1000, and the mean and standard deviation of the F1 classification score is reported for three different hyperspectral datasets.

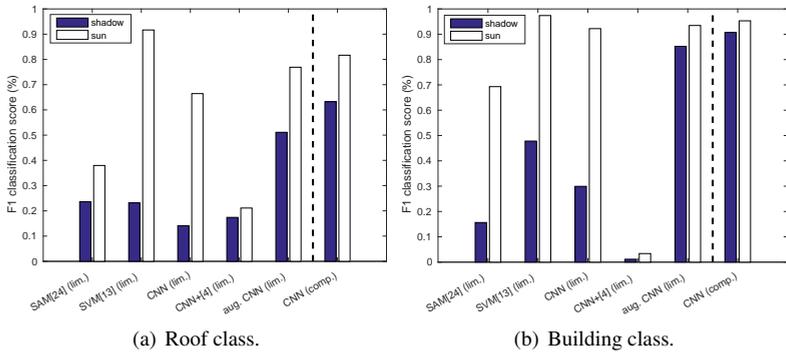


Figure 6: Spectral Relighting Augmentation Results: F1 classification score of Great Hall classes in shadow and sunlight, for several different methods, trained on limited regions and the whole image (comprehensive), for training set size 100 samples per class.

classification scores with augmentation in comparison to not using augmentation for all training set sizes and all datasets, with most of the gap between training the CNN with limited and comprehensive training sets being bridged. The results of training with the augmented data was also superior to the CNN trained on data that was projected into an illumination invariant space [4]. The SAM and SVM classifiers performed comparably on the mine face and Gualtar steps datasets, which had quite discriminative classes, but performed significantly worse on the Great Hall dataset which had more spectrally similar classes, suggesting that their use may be limited to scenarios with highly distinguishable classes. The performance of the augmented and non-augmented CNN as the amount of training samples was increased was relatively consistent. This is because the training examples were being sampled from such a small region of the image such that increasing them allowed the CNN to capture very little extra variability. Since the augmented CNN simulated the missing variability there was

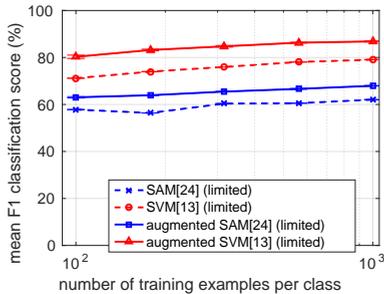


Figure 7: Spectral Relighting Augmentation Results: Evaluating the benefit of using spectral relighting augmentation for classifiers other than CNNs. SVM and SAM classifiers were trained on limited, well-illuminated regions of the Great Hall dataset with and without augmentation. The number of training samples is varied logarithmically between 100 and 1000, and the mean and standard deviation of the F1 classification score is reported.

little dependence of the classifier’s performance on the number of training examples.

Figure 6 shows the improvement that augmented training data provided specifically in the shadowed and sunlit areas for two of the Great Hall classes. In both sun and shadow, most of the gap between training the CNN with limited and comprehensive data was bridged by the augmentation, suggesting that it allowed the limited amount of training data to capture the variability as if the labels were collected from all over the image - covering areas with different geometry, incident illumination and occlusions. Also, in contrast to the SVM, which performed well on the sunlit regions for these two classes but poorly on the shadowed regions, the augmented CNNs performance in the shadow approached its performance in the sun. The CNN trained on the illumination invariant projection of the data [2] had similar performance in sunlit and shadowed regions but was significantly worse than the augmented CNN, whose performance in the shadow surpassed all other methods trained on limited data.

Figure 7 shows that the proposed spectral relighting augmentation is agnostic of the classifier. The performance of SAM and SVM classifiers are also improved upon by the data augmentation when there is limited training data available.

Despite the diffuse assumption made in (1), the results show that this is not a particularly strong constraint on the data, with each dataset possessing non-Lambertian materials. Figure 3(c) shows that there are still sporadic misclassifications in areas of shadow which are potentially due to the increased prevalence of indirect illumination which was not incorporated into the model for simplicity.

5 Conclusion

This work has proposed a method for training a CNN to be robust to several factors of variation in outdoor hyperspectral data with only a limited amount of training data. This was done through spectral relighting augmentation using an approximation of the terrestrial sunlight-diffuse skylight ratio found from the scene. Future work involves the incorporation of spatial information and training the CNN to be robust to illumination over different days.

References

- [1] Alexander Berk, Lawrence S. Bernstein, and David C. Robertson. MODTRAN: A moderate resolution model for LOWTRAN. Technical report, Spectral Sciences Inc Burlington MA, 1987.
- [2] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the Devil in the Details : Delving Deep into Convolutional Nets. In *Proceedings of the British Machine Vision Conference*, 2014.
- [3] Peter Corke, Rohan Paul, Winston Churchill, and Paul Newman. Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2085–2092. IEEE, November 2013.
- [4] Mark S Drew and Amin Yazdani Salekdeh. Multispectral Image Invariant to Illumination Colour , Strength , and Shading. *IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics*, (January):78760A–78760A, 2011.
- [5] G D Finlayson, S D Hordley, C Lu, and M S Drew. On the Removal of Shadows From Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (1):59–68, 2006.
- [6] Graham D Finlayson, Mark S Drew, and Cheng Lu. Intrinsic Images by Entropy Minimization. In *European Conference on Computer Vision*, pages 582–595, 2004.
- [7] Ola Friman, Jorgen Ahlberg, and Gustav Tolt. Illumination and shadow compensation of hyperspectral images using a digital surface model and non-linear least squares estimation. In *Image and Signal Processing for Remote Sensing XVII*, 2011.
- [8] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *Journal of Sensors*, 2015, 2015.
- [9] Alex Krizhevsky and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [10] Hyungtae Lee and Heesung Kwon. Contextual Deep CNN Based Hyperspectral Classification. *arXiv*, 1604.03519:2–4, 2016.
- [11] Hannibal H Madden. Comments on the Savitzky-Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data. *Analytical Chemistry*, 50 (9):1383–1386, 1978.
- [12] Konstantinos Makantasis, Konstantinos Karantzalos, Anastasios Doulamis, and Nikolaos Doulamis. Deep Supervised Learning for Hyperspectral Data Classification through Convolutional Neural Networks. In *IGARSS 2015. 2015 IEEE International Geoscience and Remote Sensing Symposium. Proceedings*, pages 4959–4962, 2015.
- [13] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, 2004.

- [14] Richard J. Murphy, Sildomar T. Monteiro, and Sven Schneider. Evaluating Classification Techniques for Mapping Vertical Geology Using Field-Based Hyperspectral Sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8):3066–3080, August 2012.
- [15] Sérgio M C Nascimento, Kinjiro Amano, and David H Foster. Spatial distributions of local illumination color in natural scenes. *Vision Research*, 120:39–44, 2016.
- [16] Lerrel Pinto and Abhinav Gupta. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. *arXiv preprint arXiv:1509.06825*, 2015.
- [17] Rishi Ramakrishnan. *Illumination Invariant Outdoor Perception*. PhD thesis, 2016.
- [18] Rishi Ramakrishnan, Juan Nieto, and Steve Scheduling. Shadow Compensation for Outdoor Perception. In *IEEE International Conference on Robotics and Automation*, pages 4835–4842, 2015.
- [19] Sven Schneider, Arman Melkumyan, Richard J. Murphy, and Eric Nettleton. A geological perception system for autonomous mining. *IEEE International Conference on Robotics and Automation*, pages 2986–2991, May 2012.
- [20] Max Schwarz, Hannes Schulz, and Sven Behnke. RGB-D Object Recognition and Pose Estimation based on Pre-trained Convolutional Neural Network Features. In *IEEE International Conference on Robotics and Automation*, pages 1329–1335, 2015.
- [21] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper with Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [22] Marina Trierscheid, Johannes Pellenz, Dietrich Paulus, and Dirk Balthasar. Hyperspectral Imaging for Victim Detection with Rescue Robots. In *IEEE International Workshop on Safety, Security and Rescue Robotics*, number October, pages 7–12, 2008.
- [23] Alexander Wendel and James Underwood. Self-Supervised Weed Detection in Vegetable Crops Using Ground Based Hyperspectral Imaging. In *IEEE International Conference on Robotics and Automation*, pages 5128–5135, 2016.
- [24] RH Yuhas, Alexander F H Goetz, and Joe W Boardman. Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm. *Summaries of the Third Annual JPL Airborne Geoscience Workshop, JPL Publ. 92-14*, 1: 147–149, 1992.
- [25] Wenzhi Zhao and Shihong Du. Spectral-Spatial Feature Extraction for Hyperspectral Image Classification : A Dimension Reduction and Deep Learning Approach. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4544–4554, 2016.