SAND2019-8633C

Paired neural networks for hyperspectral target detection

Dylan Z. Anderson[a], Joshua D. Zollweg[a], and Braden J. Smith[a]

[a]Sandia National Laboratories, PO Box 5800, Albuquerque, USA

ABSTRACT

Spectral matched filtering and its variants (e.g. Adaptive Coherence Estimator or ACE) rely on strong assumptions about target and background distributions. For instance, ACE assumes a Gaussian distribution of background and additive target model. In practice, natural spectral variation, due to effects such as material Bidirectional Reflectance Distribution Function, non-linear mixing with surrounding materials, or material impurities, degrade the performance of matched filter techniques and require an ever-increasing library of target templates measured under different conditions. In this work, we employ the contrastive loss function and paired neural networks to create data-driven target detectors that do not rely on strong assumptions about target and background distribution. Furthermore, by matching spectra to templates in a highly nonlinear fashion via neural networks, our target detectors exhibit improved performance and greater resiliency to natural spectral variation; this performance improvement comes with no increase in target template library size. We evaluate and compare our paired neural network detector to matched filter-based target detectors on a synthetic hyperspectral scene and the well-known Indian Pines AVIRIS hyperspectral image.

Keywords: hyperspectral, target detection, deep learning, machine learning

1. INTRODUCTION

Hyperspectral imagery (HSI) consists of hundreds of contiguous and narrow spectral bands, providing rich discriminative information for material and target detection applications.[1] These applications are typically characterized by targets of interest which are exceedingly rare: a full image consisting of millions of pixels may contain only a few dozen target pixels. Furthermore, targets often occupy only a fraction of their total pixel extent (fill-factor), with the rest of the pixel containing background clutter. These characteristics of HSI target detection have lead to the development of numerous statistical detectors, such as the matched filter, orthogonal subspace detector, mixture-tuned matched filter, kernelized matched filters, and adaptive cosine estimator (ACE).[2] In the era of machine learning and big data, these relatively simple detectors have continued to reign supreme owing to their robust and easy to use nature, minimal free parameters to select or fit, and generally good performance for a variety of targets and backgrounds. ACE in particular has enjoyed continued success repeatedly, demonstrating useful performance across a broad swath of applications.[3]

The ACE detector is derived under an additive signal model with Gaussian distributed target and background clutter.[4] Furthermore, the target and clutter are assumed to share the same covariance matrix, $\mathbf{C_b}$, which is assumed known up to a scaling factor, σ_b^2. Formally, the measured hyperspectral signal is assumed to be distributed as

$$x \sim N\left(as, \sigma_b^2 \mathbf{C_b}\right) \tag{1}$$

where s represents the known target spectra and a represents the target fill factor for the particular measured pixel, x. A generalized likelihood ratio test (GLRT) is then used to compare target present ($a > 0$) and target absent ($a = 0$) hypotheses to produce detections. Under these assumptions, the GLRT is the uniformly most powerful (UMP) detector. The "adaptive" in ACE implies that $\mathbf{C_b}$ is estimated from the data, often in a local window around the pixel under test.

Despite these fairly simplistic and rigid assumptions, ACE has repeatedly demonstrated practical success for a variety of problems. By estimating clutter statistics locally, the detector is robust to a variety of background

Further author information: (Send correspondence to Dylan Anderson)
Dylan Anderson: E-mail: dzander@sandia.gov, Telephone: 1 505 844 7013

types, including scenes containing multiple large thematic land cover types (multi-modal clutter). However, the performance of ACE and many other statistical detectors degrades when the target is not well approximated by the model assumptions. This can be driven by in-scene target morphology, collection geometry and illumination (e.g. the bi-directional reflectance distribution function of the material), errors in reflectance calibration, non-opaque materials that mix non-linearly with the background, and many other effects. These challenges are addressed by collecting a library of target spectra under as many different conditions as possible, and then applying ACE repeatedly for all or a suitable subset of the library; for sufficient library coverage it is likely that at least one target instance will satisfy the model assumptions. However, assembling such a target library is non-trivial and often requires many ground truth collection efforts to establish sufficient coverage. Furthermore, matching against a library greatly increases required computation and construction of a library must be performed for each new target material.[5,6] We refer to this as the expanding library problem.

The sub-discipline of machine learning known as deep learning has made dramatic improvements in many traditional signal processing applications. Deep learning is based on training neural networks to learn generalized representations and perform tasks using extremely large datasets.[7] Deep learning has become state of the art in many domains including natural language processing,[8] computer vision,[9] and speech recognition.[10] Recently, deep learning has been applied to hyperspectral image analysis, demonstrating impressive performance for spectral classification.[11,12] Despite this, deep learning techniques have failed to obtain widespread adoption for HSI target detection. Rarity of target materials makes training a binary classifier for target detection impractical for all but a few well characterized scenes and materials. Furthermore, the closed-set nature of standard neural network classifiers mandates at least partial re-training for detecting a new target material, further exacerbating data limitations.

This work is heavily motivated by the general success of ACE and other statistical detectors, and is developed specifically to address the expanding library problem. Our motivating question is as follows: can we design a transform, $x' = h_\theta(x)$, that when applied to a spectrum mitigates the impacts of natural spectral variability for target detection? Such a transform would integrate neatly into existing detection workflows while eliminating the need for ever-expanding libraries for each target material. In this work, we develop paired neural networks using the contrastive loss (also known as "matching networks" or "siamese networks") to construct this transform in a data-driven fashion. Matching neural networks have enjoyed state-of-the-art success in few- and one-shot learning tasks for facial recognition and biometric authentication.[13,14] This approach assumes that the mechanisms for spectral variation are common across many targets; for instance, errors resulting from apparent reflectance calibration will affect two different target materials similarly. Our framework learns to embed pixels containing targets closer to their library representations and pixels not containing targets further away, thereby non-linearly embedding spectra invariant to natural spectral variation without observing every target in every condition. After embedding, any standard statistical detector (including ACE) can be applied to perform target detection. Since this framework learns an embedding space suitable for detection rather than learning to directly perform detection, it can readily generalize to new targets never seen during training with just a single target prototype.

2. CONTRASTIVE LOSS AND PAIRED NEURAL NETWORKS

The hyperspectral target detection problem can be cast as follows. Given a small set (possibly as few as one) spectral prototypes of a target material, compute the distance between the prototypes and the pixels in the image. Pixels which have a small distance to prototypes likely contain the target material and likewise pixels with a large distance likely do not contain the target material. To mitigate the expanding library problem, we would like to find some transform, denoted as h_θ, that, when applied to our prototype and in-scene spectra, reduces or eliminates the effects of natural spectral variation on measured distances. This target detection framework is shown graphically in Figure 1.

More precisely, our objective is to learn an embedding function $h_\theta(x) : \mathbb{R}^B \rightarrow \mathbb{R}^K$ which maps length B spectra onto length K vectors. We denote the euclidean distance between two embedded spectra as

$$D_\theta = \|h_\theta(p_t) - h_\theta(x_i)\|_2 \tag{2}$$

where p_t represents the spectral prototype for target material t and x_i represents the spectrum of pixel i. Note that it is possible to use other distances, such as the L1 distance, provided that they are differentiable with

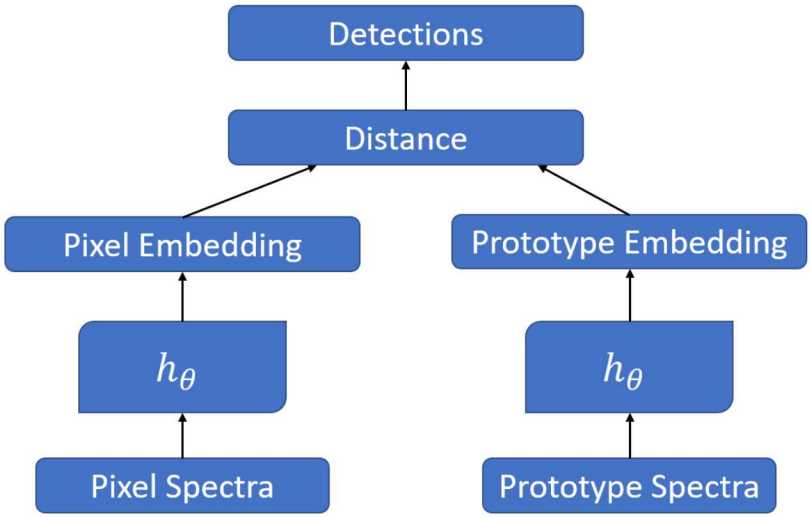

Figure 1: Target detection framework. The embedding transform, h_θ, applied to pixel and prototype spectra can take on many different forms such as linear transforms or deep neural networks.

respect to θ. For the embedding function to be suitable for hyperspectral target detection, we desire that D_θ be small if x_i contains the target material and large if it does not contain the target material. These properties are codified in the contrastive loss, given by[15]

$$\mathcal{L}\left(p_t, x_i\right) = y\frac{1}{2}D_\theta^2 + (1-y)\frac{1}{2}\{max(0, m - D_\theta)\}^2 \tag{3}$$

where y is an indicator variable, with $y = 1$ when x_i contains material t and $y = 0$ when it does not. The parameter $m > 0$ defines the margin, above which non-target (prototype, pixel) pairs are deemed sufficiently far apart and do not further contribute to the loss. Contrary to loss functions traditionally used for classification or regression, such as cross entropy or mean squared error, the contrastive loss is defined for pairs of spectra and a target indicator rather than a single input and corresponding label.

The contrastive loss does not dictate anything about the functional form of h_θ and it can take on many different forms ranging from linear transforms to non-linear deep neural networks. To address the expanding library problem, we specifically desire a model that embeds robustly to non-linear natural spectral variation. Given this, along with the general practical success of ACE, we choose h_θ to be relatively simple fully-connected neural networks joined by Rectified Linear Unit (ReLU) activation functions. We then learn appropriate parameters θ by training our model with respect to the contrastive loss function.

It is often more natural to interpret a score of target similarity rather than distance. To convert embedding distances to similarities, we employ the following transform:

$$S_\theta = \frac{1}{1 + D_\theta} \tag{4}$$

There are many possible forms to convert a distance into a similarity, but we use this simple transform as it does not introduce any additional parameters (such as σ for a radial basis function) and is readily invertible back to distance. To make detections, we first embed our prototypes using our trained network. Then when a new scene is collected, we simply have to embed the scene pixels and measure the distance (similarity) to the desired prototypes. Since our network has been trained to embed robustly to natural spectral variation, our prototype library requires far fewer examples of each target material. Threshold euclidean distance is equivalent to applying a spectral matched filter in the network embedding space to make detections.

3. EXPERIMENTS

Our experimental evaluation is split into two sections. In the first section we evaluate pairwise networks using the well-known Indian Pines hyperspectral dataset.[16] In the second section we evaluate our approach using a synthetic DIRSIG scene[17] designed to reflect common challenges associated with natural spectral variation.

3.1 Indian Pines

We first evaluate the paired network approach using the Indian Pines dataset. Indian Pines is a 145×145 hyperspectral image of Northwest Indiana. The scene is composed approximately two-thirds as agricultural fields and the remaining as forest and other native vegetation. The data, collected by AVIRIS, contains 16 different classes and 224 spectral bands from $0.4 - 2.5\mu m$, with 24 water absorption bands discarded. In utilizing Indian Pines to evaluate target detectors, we have selected the "corn-notill" class as the target of interest. The mean-spectral response, crop-classes, and target truth are shown in Figure 2.

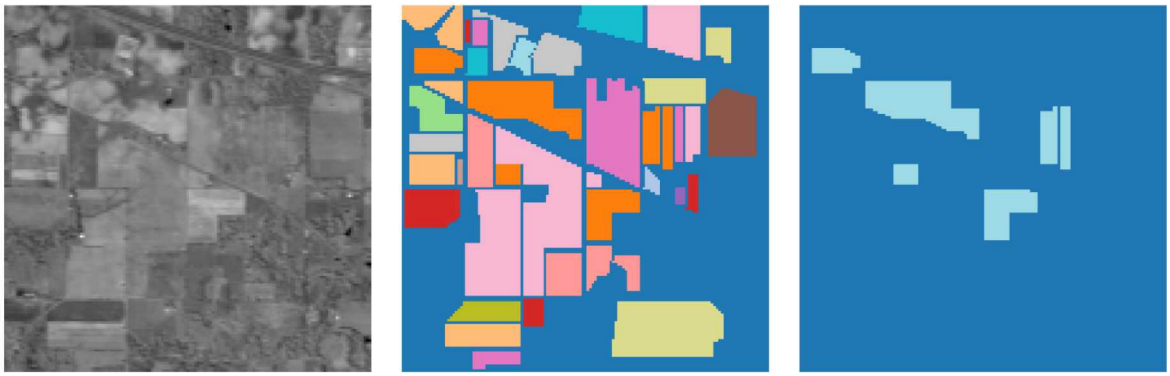

Figure 2: Indian Pines mean spectral response (left), crop class truth (center), and corn-notill target truth (right).

The Indian Pines dataset is split into 60% train, 20% validation, and 20% test. We utilize the mean spectrum of target pixels from the train set as a target prototype. We baseline our approach against ACE, for which we estimate background statistics using the non-target pixels from the train set. Our strategy for generating the (pixel, prototype, indicator) training tuples is as follows. For each pixel in the train set, we sample 10 target pixels and 10 non-target pixels. Tuples are constructed using the original pixel as the prototype, the sampled pixels as tuple pixels, and the indicator set to 1 if pixel and prototype *both* contain or do not contain the target. This results in 20 tuples per training pixel. We employ a simple 150-100-100-50-20 (5 layer) fully connected neural network with ReLU activations between layers. The model is trained for 10 epochs using the Adam optimizer, with a contrastive loss margin parameter of 2. After training, we measure distances between test pixels and the mean spectrum prototype. To evaluate target detection performance we utilize the probability of detection for 5% constant false alarm rate. Results for our paired network approach and ACE are shown in Figure 3.

While our approach handily outperforms ACE in this test, these results are a greater indication that the paired network framework can perform on-par with traditional supervised binary classifiers than that it provides superior target detection performance. This test does not feature many of the hallmarks of target detection applications: target pixels are numerous for training, the labeled classes in Indian Pines represent large, many-pixel crop fields resulting in "pure" target pixels, and the dataset consists of a single collection, mitigating spectral variation induced by changes in atmosphere, simplifying assumptions in reflectance calibration, etc. Furthermore, the target prototype is generated directly from in-scene data which is unrealistic in many practical collection scenarios. Indian Pines is commonly used in hyperspectral classification studies since per-pixel ground truthing is available, but it is less appropriate for evaluating target detection algorithms.

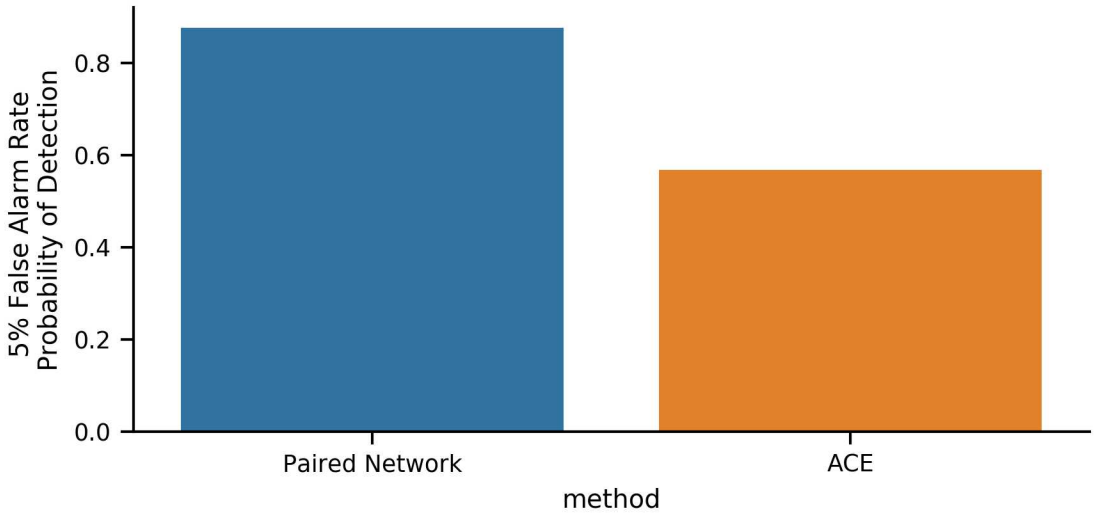

Figure 3: Target detection performance for the Indian Pines experiment.

3.2 DIRSIG Megascene

To address the shortcomings of the Indian Pines experiment and to better evaluate the effects of spectral variability, we developed a synthetic dataset based on DIRSIG Megascene.[17] Megascene covers over half a square mile designed to represent an area of Northeast Rochester, NY. The scene is composed of numerous man-made structures (houses, large buildings, roads, vehicles) as well as highly vegetative areas composed of various types of trees and grasses. A pseudo-color quicklook image of megascene in shown in Figure 4. Imagery were simulated using an AVIRIS-like sensor with 211 bands from 0.4 to 2.5 μm. Imagery were captured as a single snapshot from a nadir look geometry centered over the scene from a sensor elevation of 4 km, resulting in an effective ground sample distance of 1 m. In total, nine different images were simulated for combinations of three different MODTRAN-based atmospheres and three different times of day (1200, 1430, 1545). Atmospheres considered were Mid-Latitude Summer (MLS), Sub-Artic Summer (SAS), and Tropical (TROP). For each image, a radiance to reflectance conversion factor was applied:

$$K = \frac{\pi}{E_{TOA} t_1 t_2 cos\left(\theta_{sun}\right)} \tag{5}$$

where E_{TOA} represents top of atmosphere irradiance, t_1 is sun to ground transmission, t_2 is ground to sensor transmission, and θ_{sun} is the sun angle from nadir. E_{TOA} was provided directly from DIRSIG and atmospheric transmissions were derived utilizing MODTRAN with the corresponding atmospheric profile. This reflectance calibration makes the simplifying assumption of Lambertian surfaces for all materials in the scene. While this is not perfectly accurate, this assumption is common in realistic collection scenarios and is an additional source of spectral variation. These simulations for varying atmospheres and times of day coupled with realistic but imperfect reflectance calibration yield images that exhibit types of natural spectral variation common in real world collections. Alongside the simulated imagery, a spectral library consisting of a single spectral prototype for each material in the scene (113 unique materials) was also produced. Prototypes were generated using DIRSIG with the same sensor model from nadir at 200 m elevation to image a single "pure" pixel consisting of the material of interest under MLS-1200 conditions. The same reflectance calibration was applied to the library. This library represents realistic constraints of assembling a practical detection library: it is often unfeasible to collect prototypes for targets of interest under all possible collection conditions.

Throughout the scene we inserted additional targets of interest made of three different materials drawn from the Nonconventional Exploitation Factors Data System (NEFDS): two green paints (0887UUUPNT and 0688UUUPNT) and a yellow paint (0872UUUPNT). These materials, which we refer to as green_paint_1, green_paint_2,

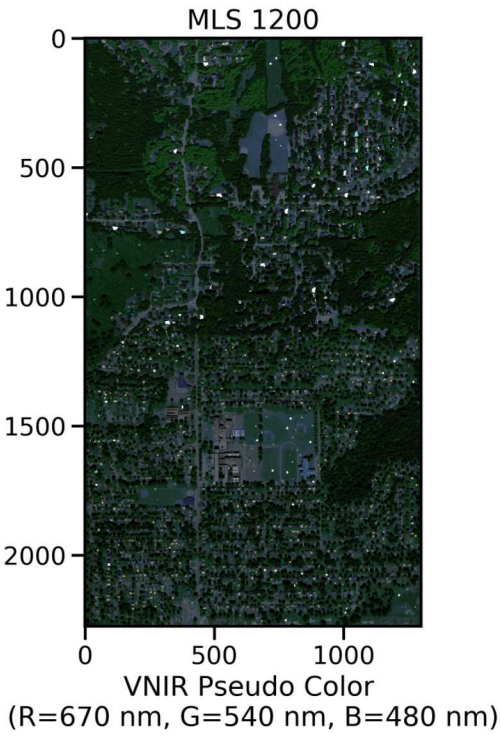

Figure 4: Pseudo color quicklook of the MLS-1200 megascene image.

and yellow_paint, include hemispherical reflectance and bi-directional reflectance distribution functions in the Maxwell-Beard form. The spectral prototypes for these material in our spectral library are shown in Figure 5. In each image, 125 disks of varying radii from 0.1 to 4 m for each target material were inserted in randomized locations. Yellow paint targets were inserted only on the righ half of each image. A bump map was applied to each target disk to further increase spectral variability. Target disks as seen by the sensor as well as a high-fidelity sub-pixel rendering of a target randomized underneath foliage is shown in Figure 6.

Only the left-half of the MLS-1200 render and the material prototype library was used for training and validation. The right half of all nine renders was used as a test set, utilizing the same probability of detection for 5% constant false alarm rate as in the Indian Pines experiment for evaluation. We evaluate performance against two different targets: green_paint_1, for which we explicitly inserted green_paint_2 as an additional confuser and yellow_paint, which never occurs in the in-scene training or validation data. A defining feature in many target detection applications is sub-pixel targets. We divide our performance metrics across three different bins for target fill fraction to ensure performance both sub- and pure-target pixels: low (up to 0.25 fill fraction), medium (0.25 to 0.75 fill fraction), and high (above 0.75 pixel fill fraction). We again utilize ACE as a performance baseline. We utilize the material prototype library as target templates, and estimate background statistics in adaptive mode. We use a guard band of 11 pixels, corresponding to our maximum estimated target size, and sweep over background window sizes from 51 to 201. We report performance utilizing the best performing windows for each individual test render.

In contrast to the Indian Pines experiment, we utilize all materials (not just the targets of interest) for training. We employ the following strategy to generate (prototype, pixel, indicator) tuples for training. For each target material t present in the training set, utilize the corresponding library prototype p_t as the tuple prototype. Sample 3,000 pixels that contain the material t (fill fraction > 0). Given the rarity of many target materials it is likely that the total number of pixels that contain t is less than 3,000, in which case we simply sample all the available target containing pixels. For these pairs, set the indicator $y = 1$ and bundle into tuples $(p_t, x_i, y = 1)$. Next, sample an equal number of pixels that do not contain target materials t (fill fraction $= 0$)

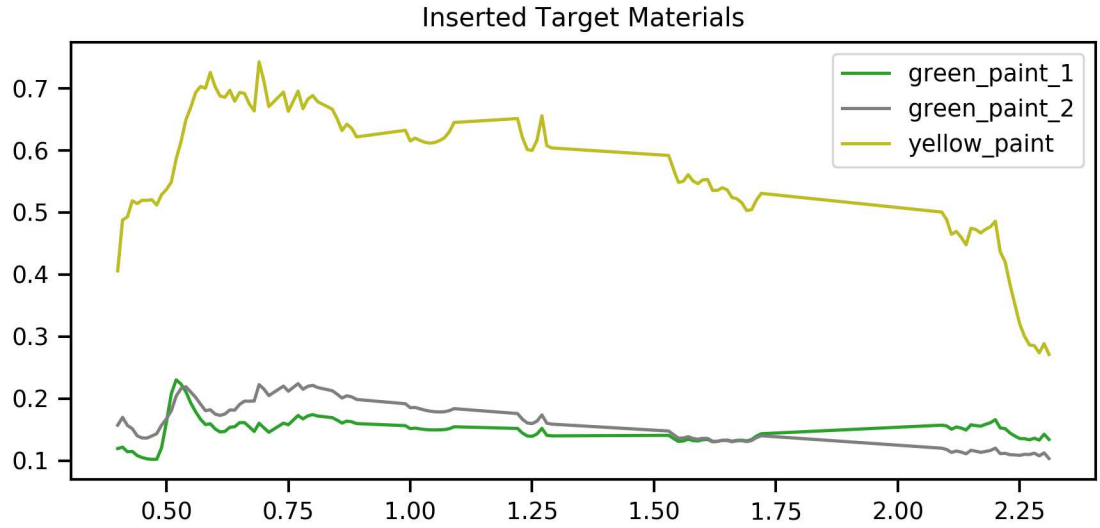

Figure 5: Spectral prototypes for inserted target materials.

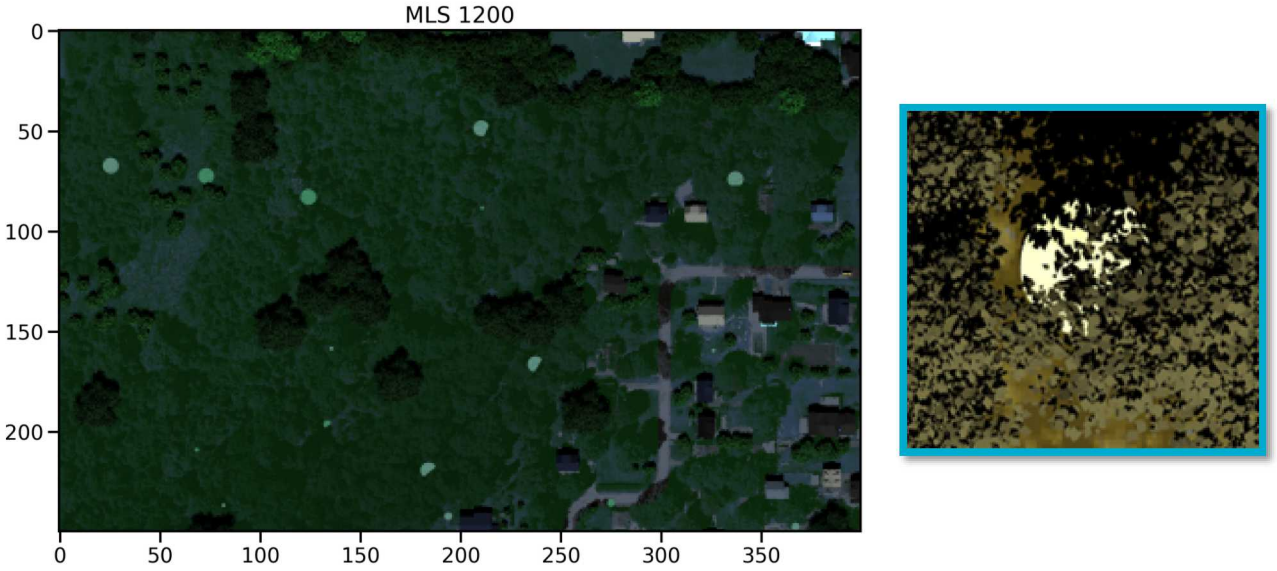

Figure 6: Disk targets inserted throughout the scene. On left, green_paint_1 and green_paint_2 disks inserted in the MLS-1200 simulation. On right, a high fidelity (sub-pixel) rendering of a target that was randomized beneath foliage.

and bundle into tuples $(p_t, x_i, y = 0)$. This results in up to 6,000 training tuples per material in the training set. This sampling strategy leaves a considerable number of non-matched pairs unused for each material. To improve sample diversity, we resample the training tuples every 100 epochs. Since for most materials there are fewer than 3,000 material-containing pixels in the training set, this results in many of the same matched tuples with a rotating set of non-matched tuples. To prevent leakage of pixels from train into the validation set, we spatially partition the dataset into train and validation. The bottom 500 rows of the left side of the MLS-1200 image are used as validation, with the remainder of the left side used for training. Note that since tuples are sampled only for materials occurring in the training set, it is possible for entirely new materials to be present in validation or

test not seen in training, such as the yellow_paint which occurs only in the test set and we use for evaluation.

We again employ a simple fully connected neural network with ReLU activations and layer sizes 150-150-100-100-50-25 (see Figure 7). We train our model for a total of 10,000 epochs utilizing the Adam optimizer with a contrastive loss margin parameter of 175. After training, distances between test set embeddings from all nine renderings and embedded green_paint_1 and yellow_paint prototypes (one per target) are thresholded to make detections.

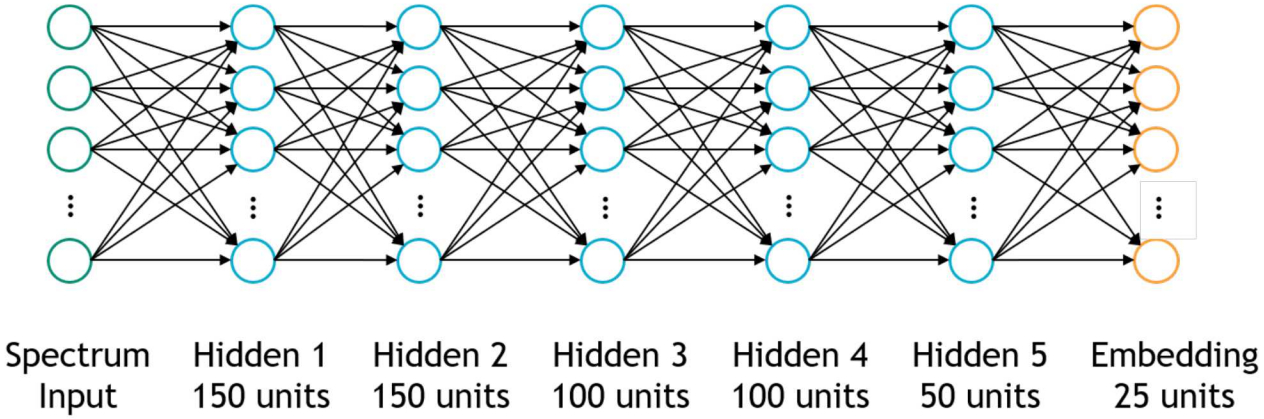

Figure 7: Fully connected network architecture utilized for Megascene experiment. ReLU activations are utilized between layers.

Figure 8 shows the target detection performance for just the right side of MLS-1200 as evaluation. For green_paint_1, ACE and paired networks performance are comparable, with paired networks only slightly edging out ACE. It is expected for ACE to show strong performance on the MLS-1200 test case; the target library prototype is well aligned to the MLS-1200 render eliminating a great deal of natural spectral variation. Surprisingly, the performance gap for yellow_paint is even more favorable to paired networks, though ACE still provides solid performance. It is noteworthy that paired networks achieves this performance level despite having only the library prototype for yellow_paint and no in scene observations. Figure 9 shows target detection performance averaged over all nine atmosphere and time of day test renders. Compared over the full range of test spectral variability, the performance gap between paired networks and ACE widens considerably. The paired network approach maintains its high performance level over all collection conditions and target fill fractions with just a single library prototype for each target, demonstrating robustness to natural spectral variation.

4. DISCUSSION

While the paired network approach provides large performance improvements in our experiments, it does require additional training data compared to ACE. The tuple sampling process requires material composition knowledge for many in-scene pixels. However, the framework does not require that these tuples come from only (likely rare) targets of interest, as demonstrated by the yellow_paint prediction performance despite having no training examples of that material. The number of tuples for training can be greatly bolstered using ubiquitous and often easy to label background clutter materials such as types of vegetation, asphalt or other anthropogenic artifacts, soil or geologic types, etc. We hypothesize that for the paired networks framework to generalize well to new targets, it does not have to be trained explicitly with those targets but it should be trained on examples that exhibit the same natural spectral variation phenomenology. In other words, the *mechanisms* for variation expected for targets of interest should be captured in the training examples. Realistically, it is likely far easier to generate many of these tuples for a variety of common targets than to generate an intricate detection library for a single target of interest (which does not generalize to new target materials).

In our experiments, we utilize fairly simplistic fully-connected neural network architectures. In order to compensate for natural spectral variability, we desire the ability to represent highly non-linear transforms as

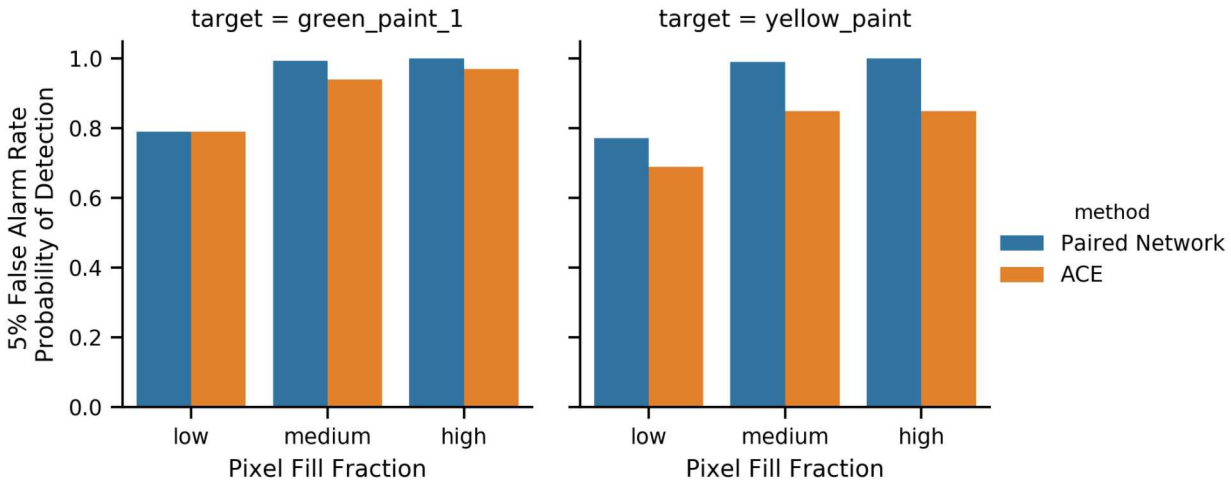

Figure 8: Target detection performance against only the MLS-1200 right half test set. The paired network approach provides a slight performance edge over ACE.

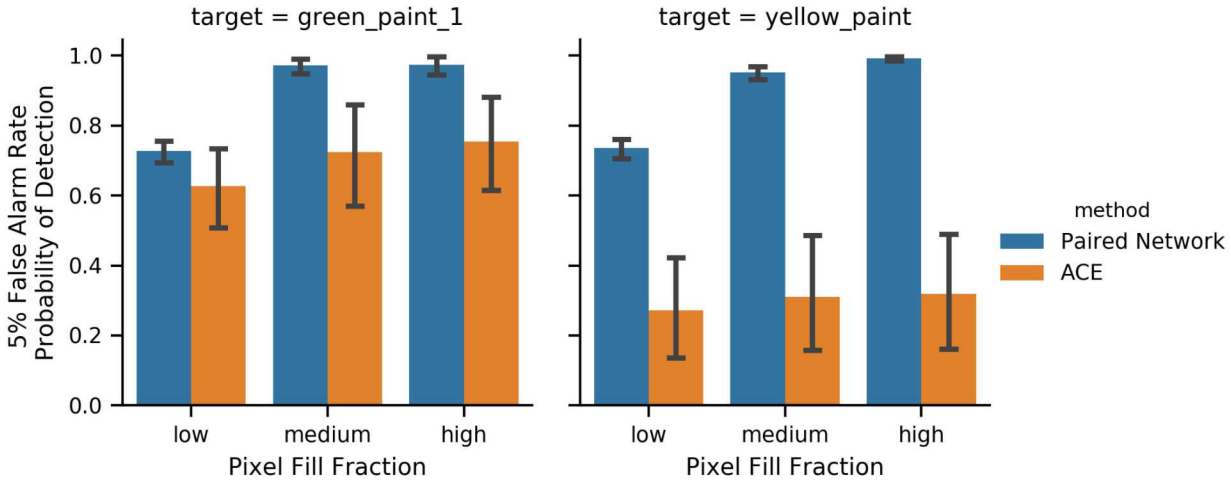

Figure 9: Target detection performance averaged over all nine test set conditions. The paired network approach provides significant improvement over ACE since it is robust to natural spectral variation.

provided by multi-layer neural networks. However, the importance of specific spectral band values as opposed to shapes in spectroscopy makes a fully connected architecture more suitable than a convolutional network for purely spectral input. Had we utilized spatial-spectral information, a convolutional network may have been appropriate. Additionally, by reducing the size of each successive layer in our architectures, we encode a low-rank assumption on the hyperspectral data. A low rank assumption is common in hyperspectral applications to capture the dominant linear mixing of a few select endmembers. If we did not expect this low-rank assumption to hold, we could just as easily increase layer sizes so that the embedding space maintained the number of spectral bands. Finally, we observe that statistical detectors still enjoy widespread success for hyperspectral target detection and therefore highly sophisticated networks are likely not needed for most realistic scenarios.

Training paired networks does require additional computational time: our megascene experiment ran for 30 hours on an Nvidia Titan V GPU to complete the 10,000 epochs. However, after training the model could be

used to perform inference and make detections in real time. ACE does not require training but the adaptive nature necessitates continually re-estimating background spectral covariance around each in-scene pixel to make detections. The medium sized windows used by ACE in the megascene experiment required 7 hours per render to compute (approximately 60 hours total for all nine renders). This is longer than the training time required for paired networks, and would be incurred again for additional scenes or collections. Furthermore, the covariance matrix estimation and inverse as required by ACE scales cubicly in the number of spectral bands; for the next generation of higher spatial-spectral resolution sensors this compute cost will grow dramatically. These factors offset the required training time for the paired network approach.

5. CONCLUSION

In this paper, we present paired neural networks and contrastive loss for hyperspectral target detection. We develop this framework specifically to address the expanding library problem and fragility of classical statistical detectors such as ACE to natural spectral variability. The contrastive loss quantifies robustness to this spectral variation, and enables using paired neural networks to learn a transform that mitigates the performance degradation on target detection. Since this framework generates a transform, the trained model can fit seamlessly into existing target detection workflows and can be paired with any statistical detector such as spectral matched filtering or adaptive cosine estimation for final detections. In our experiments, we demonstrate greatly improved detection performance over ACE in realistic scenarios with spectral variation. The one-shot learning mechanism of paired neural networks enables making detections of entirely new target materials with just a single target prototype and no retraining of the network.

ACKNOWLEDGMENTS

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] Nasrabadi, N. M., "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Processing Magazine* **31**(1), 34–44 (2013).

[2] Manolakis, D., Marden, D., Shaw, G. A., et al., "Hyperspectral image processing for automatic target detection applications," *Lincoln laboratory journal* **14**(1), 79–116 (2003).

[3] Manolakis, D., Pieper, M., Truslow, E., Cooley, T., Brueggeman, M., and Lipson, S., "The remarkable success of adaptive cosine estimator in hyperspectral target detection," in [*Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*], **8743**, International Society for Optics and Photonics (2013).

[4] Kraut, S., Scharf, L. L., and McWhorter, L. T., "Adaptive subspace detectors," *IEEE Transactions on signal processing* **49**(1), 1–16 (2001).

[5] Healey, G. and Slater, D., "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Transactions on Geoscience and Remote Sensing* **37**(6), 2706–2717 (1999).

[6] Ientilucci, E. J. and Bajorski, P., "Stochastic modeling of physically derived signature spaces," *Journal of applied remote sensing* **2**(1), 023532 (2008).

[7] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *nature* **521**(7553), 436 (2015).

[8] Sutskever, I., Vinyals, O., and Le, Q. V., "Sequence to sequence learning with neural networks," in [*Advances in neural information processing systems*], 3104–3112 (2014).

[9] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).

[10] Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Kingsbury, B., et al., "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal processing magazine* **29** (2012).

[11] Chen, Y., Jiang, H., Li, C., Jia, X., and Ghamisi, P., "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing* **54**, 6232–6251 (Oct 2016).

[12] Petersson, H., Gustafsson, D., and Bergstrom, D., "Hyperspectral image analysis using deep learning — a review," in [*2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*], 1–6 (Dec 2016).

[13] Koch, G., Zemel, R., and Salakhutdinov, R., "Siamese neural networks for one-shot image recognition," in [*ICML deep learning workshop*], **2** (2015).

[14] Vinyals, O., Blundell, C., Lillicrap, T., kavukcuoglu, k., and Wierstra, D., "Matching networks for one shot learning," in [*Advances in Neural Information Processing Systems 29*], Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., eds., 3630–3638, Curran Associates, Inc. (2016).

[15] Hadsell, R., Chopra, S., and LeCun, Y., "Dimensionality reduction by learning an invariant mapping," in [*2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*], **2**, 1735–1742, IEEE (2006).

[16] Baumgardner, M. F., Biehl, L. L., and Landgrebe, D. A., "220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3," (Sep 2015).

[17] Ientilucci, E. J. and Brown, S. D., "Advances in wide-area hyperspectral image simulation," in [*Targets and Backgrounds IX: Characterization and Representation*], **5075**, 110–121, International Society for Optics and Photonics (2003).