RAIC Embeddings for Geospatial Imagery Categorization

Brian Goodwin, PhD
CTO, RAIC Labs

Abstract

RAIC is a human-AI interface technology designed to enhance collaboration between analysts and AI foundation models, enabling high-accuracy inference, low-latency feedback, and efficient model updates or retraining. By integrating unsupervised learning with a human-in-the-loop approach, RAIC facilitates the development of robust computer vision models through zero-shot learning and instance/object retrieval across massive imagery datasets. This framework empowers users to build monitoring applications, continuously assess model performance, experiment, and discover. We evaluated RAIC's performance on two established remote sensing datasets: PatternNet and EuroSAT. RAIC was exposed to a broad range of geospatial imagery during pretraining, providing a flexible embedding space suitable for various downstream computer vision tasks. For this study, we use a simple logistic regression model to evaluate the predictive power of RAIC embeddings. Our results show that RAIC can achieve state-of-the-art accuracies using only a fraction of each dataset compared to typical 90/10 train/test splits. With substantially reduced training requirements, we surpass previously published accuracies (0.967 and 0.937, respectively) on each dataset using considerably less labeled data. Specifically, these SOTA accuracies were surpassed for both PatternNet and EuroSAT using only 760 samples (2.5% of dataset) and 1000 samples (3.7% of dataset), respectively. Our experiments and t-SNE visualizations confirm that the learned embedding space is intrinsically separable, indicating RAIC's ability to represent dense visual features effectively. Our findings also show that RAIC alleviates the necessity for deep models as well as GPU hardware for many analytics tasks, thus lowering computational overhead, cost, and time in remote sensing applications. Notably, while training CNN models on these datasets requires a GPU, our models train in just 0.2 seconds on a CPU.

1 Introduction

Recent advances in deep neural networks (DNNs) have revolutionized computer vision tasks by providing powerful feature representations (i.e., embeddings) of imagery [HZRS16, CDLSB18, DBK+21, SZ15]. In the computer vision community, pretrained feature extractors such as convolutional neural networks (CNNs) and Vision Transformer (ViT) networks have enabled transfer learning to downstream tasks with reduced training data and computational costs. However, training large models often requires considerable hardware resources (e.g., large GPU clusters) and large labeled datasets, which can be both expensive and time-consuming.

In natural language processing (NLP), large language models (LLMs) have demonstrated remarkable generalization, enabling users to avoid retraining models from scratch and instead employ techniques like Retrieval-Augmented Generation (RAG) or prompt engineering to adapt the model to new tasks. This concept of "train once, use repeatedly" has significantly lowered the barrier to utilizing powerful language representations.

Inspired by these developments, we introduce a similar paradigm for geospatial image analytics using a pretrained embedding engine called RAIC. RAIC is trained on a broad set of remote sensing imagery, including multispectral data, yielding a compact representation that captures diverse spatial and spectral features [CJL+16]. By leveraging these embeddings, we can train shallow models such as logistic regression or k-Nearest Neighbors (KNN) on top of RAIC representations, thus removing the need for GPU-based training when dealing with new tasks.

Research advancements in pretraining have shown that shallow models can achieve competitive performance on image classification tasks when trained on embeddings. Many pretraining methods have emerged to create powerful embeddings, such as SimCLR [CKNH20], SwAV [CMM+20], MoCo [HFW+20], and BYOL [GSA+20] to name a few. However, these methods are often tailored to natural images and may not generalize well to remote sensing data. RAIC, on the other hand, is designed to be a general-purpose, adaptable engine, which suits a wide range of remote sensing applications. Similar to other appraoches to task specific pretraining like change detection [ZFY+17, CDLSB18, CS20], RAIC extends to a broader set of tasks, including change detection, classification, retrieval, and anomaly detection.

In this paper, we focus on two popular remote sensing benchmark datasets: PatternNet [ZLN18] and EuroSAT [Sha] (Figure 3), evaluating how well RAIC embeddings can be used to achieve or surpass state-of-the-art (SOTA) results. Prior studies have shown that models like ResNet50 can achieve 0.967 accuracy on PatternNet [ZLN18], while custom CNN approaches can reach 0.937 on EuroSAT [Sha, Art21]. We demonstrate that using as few as 20 samples per class on PatternNet and 100 per class on EuroSAT, a simple logistic regression model trained on RAIC embeddings outperforms these SOTA accuracies.

2 Methods

2.1 Datasets

RAIC was evaluated on two remote sensing datasets: PatternNet and EuroSAT. These datasets are commonly used in the remote sensing community for benchmarking classification algorithms.

PatternNet is a high-resolution remote sensing dataset consisting of 30,400 images from Google Earth covering 38 classes of man-made structures, such as parking lots, tennis courts, and ports [ZLN18]. Each image is 256×256 pixels in RGB.

EuroSAT is a dataset of 27,000 sentinel-2 satellite images covering 10 classes of land uses (e.g., residential, industrial, annual crop) [Sha]. Each image contains multiple spectral bands, but standard approaches often use the 3-band RGB subset [Art21].

2.2 RAIC Embeddings

Our RAIC embedding engine was pretrained using a fully unsupervised protocol on a large corpus of geospatial imagery covering a variety of sensors and spectral bands. The training objective was designed to create a condensed representation of the imagery, capturing salient spatial and spectral features across diverse geospatial domains. By doing so, RAIC becomes a general-purpose embedding engine capable of processing both RGB and multispectral inputs.

2.3 Experimental Setup

For each dataset, we sampled a small subset of images per class, starting at 5 images and incrementing by 5 images, up to 200 images. Each sampled image was passed through RAIC, yielding a fixed-dimensional embedding vector. A logistic regression model was then trained on these embeddings to classify images into their respective classes. We used a 50% hold-out set, randomly sampled, for evaluating model accuracy. Each sampling procedure was repeated 20 times, and the results were averaged.

We compare our accuracy values to the reported SOTA on each dataset. **PatternNet SOTA** (≈ 0.967), was obtained by a ResNet50-based approach as reported in [ZLN18]. Our research uncovered a **EuroSAT SOTA** (≈ 0.937), reported from a custom CNN [Sha], which was also validated by other attempts with ResNet50 (≈ 0.919) [Art21].

3 Results

Figure 1 (left) shows the accuracy on PatternNet as the number of training samples per class increases. Even with only 20 images per class (only 2.5% of PatternNet's full dataset), the accuracy exceeds the reported SOTA of 0.967. Beyond 50 samples per class, the logistic regression model on RAIC embeddings plateaus near 0.99, which significantly surpasses the deep model SOTA.

Similarly, Figure 2 (left) shows that using just 100 samples per class (3.7% of EuroSAT), we surpass the 0.937 SOTA. The t-SNE plots (Figures 1 and 2, right) visualize how RAIC embeddings naturally

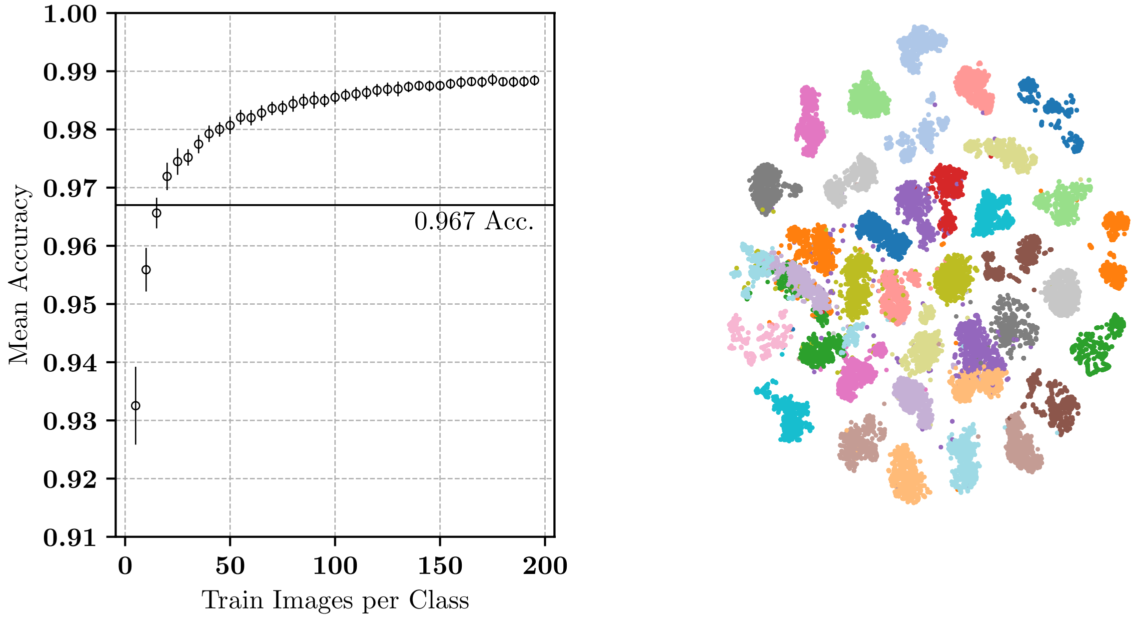

Figure 1: Left: Mean accuracy on PatternNet vs. train images per class (using a logistic regression model). Right: t-SNE visualization of RAIC embeddings for PatternNet dataset.

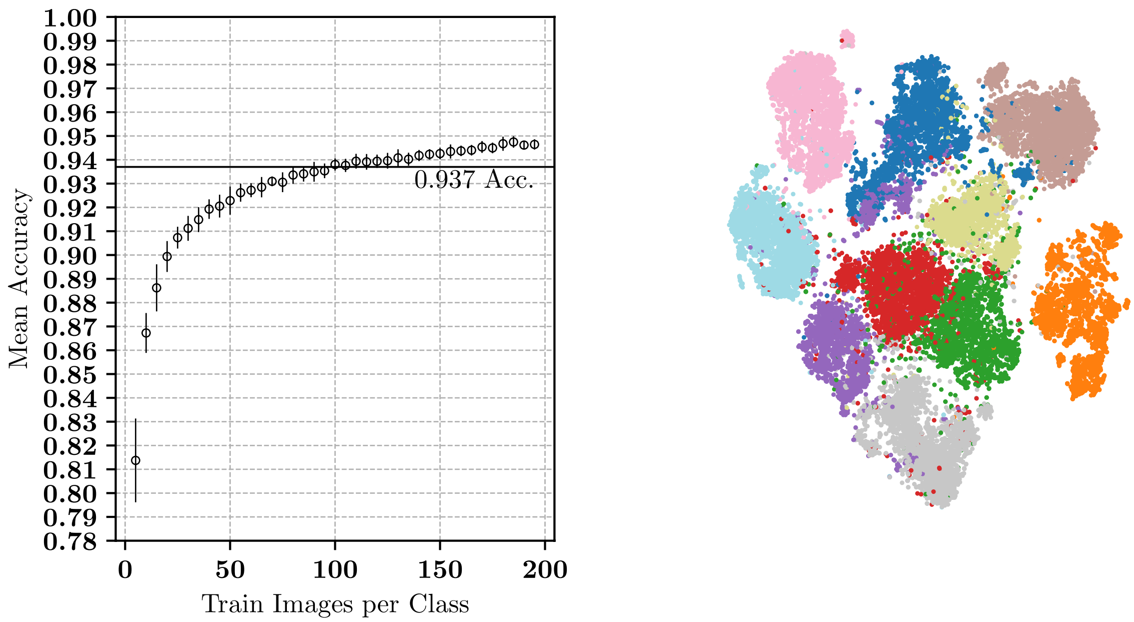

Figure 2: Left: Mean accuracy on EuroSAT vs. train images per class (using a logistic regression model). Right: t-SNE visualization of RAIC embeddings for EuroSAT dataset.

cluster different classes without specialized training on these datasets, demonstrating significant inherent separability.

4 Discussion

The RAIC embedding engine offers several advantages for geospatial analytics. Firstly, it significantly reduces training overhead by providing rich, pre-computed features, allowing classification tasks to

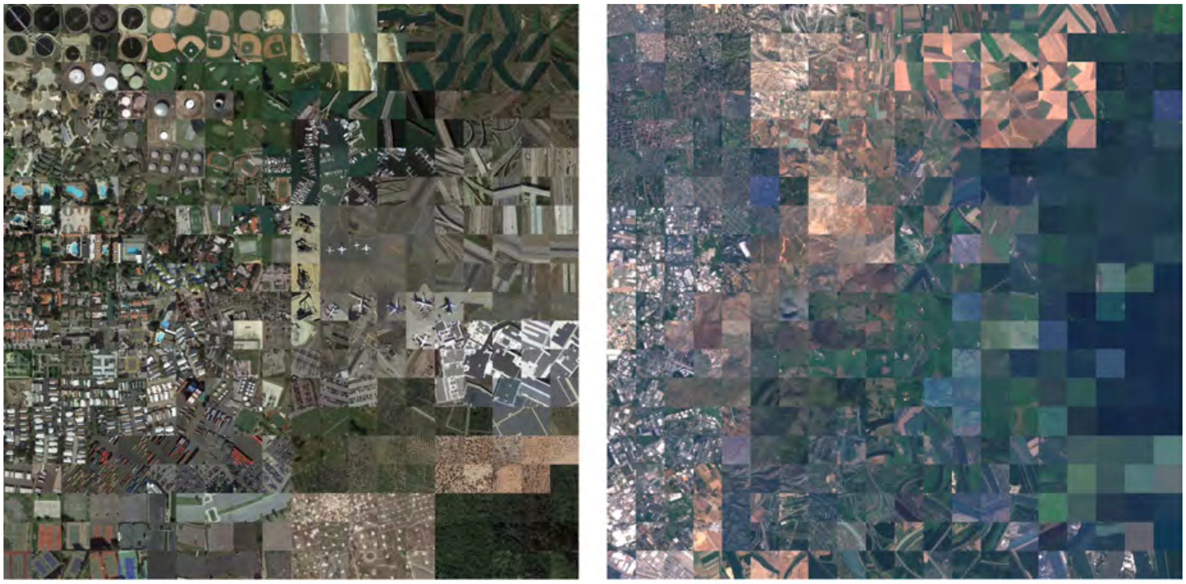

Figure 3: t-SNE grid visualizations of RAIC embeddings for 400 randomly sampled images from PatternNet (left) and EuroSAT (right).

be handled by shallow models such as logistic regression or k-NN that do not require GPU resources, thereby greatly lowering the computational barrier. RAIC embeddings enable training times of just 0.2 seconds on a CPU. Secondly, RAIC demonstrates flexibility across data domains, having been trained on a wide spectrum of remote sensing data, including hyperspectral and multispectral bands, which enables it to generalize across different sensor modalities. This is particularly beneficial for niche remote sensing applications that would otherwise require extensive retraining of a large network. Additionally, RAIC draws an analogy to large language models by facilitating powerful vision embeddings that can be adapted to numerous tasks via simple or even zero-shot approaches. Just as techniques like Retrieval-Augmented Generation (RAG) or prompt engineering can repurpose LLMs in the text domain, RAIC embeddings can be repurposed for image-based search, anomaly detection, sensor drift analysis, scene categorization, or change detection with minimal compute overhead. Lastly, our results suggest that even k-NN classifiers on RAIC embeddings can achieve high accuracy, indicating a strong separation of classes in this latent space, which can also extend to unsupervised tasks like clustering or anomaly detection.

By enabling high-accuracy classification with only a fraction of the labeled data, RAIC demonstrates efficacy, underscoring its value for remote sensing—a field where labeled data collection is particularly challenging. The success of RAIC on PatternNet and EuroSAT highlights the possibility of leveraging a single, robust embedding engine to solve a wide array of geospatial computer vision problems.

References

[Art21] Artemisart. Eurosat-image-classification github repository. https://github.com/artemisart/EuroSAT-image-classification, 2021.

[CDLSB18] Remi Caye Daudt, Bertrand Le Saux, and Alexandre Boulch. Fully convolutional siamese networks for change detection. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 4063–4067, 2018.

[CJL+16] Yun Chen, Heng-Chao Jiang, Chun Li, Xian Jia, and Pedram Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016.

[CKNH20] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proc. 37th Int. Conf. on Machine Learning (ICML)*, pages 1597–1607, 2020.

[CMM+20] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 9912–9924, 2020.

[CS20] Hao Chen and Zhenwei Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10):1662, 2020.

[DBK+21] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *Int. Conf. on Learning Representations (ICLR)*, 2021.

[GSA+20] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Remi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2020.

[HFW+20] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9729–9738, 2020.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[Sha] Sachin Sharma. Hands-on experience on achieving state-of-the-art results on classifying eurosat satellite images. https://sachinsharma9780.medium.com/hands-on-experience-on-achieving-state-of-the-art-results-on-classifying-eurosat-satellite Accessed: 2020-12-01.

[SZ15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Int. Conf. on Learning Representations (ICLR)*, 2015.

[ZFY+17] Yang Zhan, Kun Fu, Menglong Yan, Xian Sun, Hongqi Wang, and Xiaosong Qiu. Change detection based on deep siamese convolutional network for optical aerial images. *IEEE Geoscience and Remote Sensing Letters*, 14(10):1845–1849, 2017.

[ZLN18] B. Zhou, L. Liu, and S. Newsam. Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:197–209, 2018.