Self-supervised Contrastive Learning for Wildfire Detection: Utility and Limitations

Jae Won Choi[1], Nick LaHaye[2], Yuzhou Chen[3], Hugo Lee[2], Yulia R. Gel[4,5]

[1]Department of Computer Science, University of Texas at Dallas
[2]Jet Propulsion Laboratory, California Institute of Technology
[3]Department of Computer and Information Sciences, Temple University
[4]Department of Mathematical Sciences, University of Texas at Dallas
[5]National Science Foundation

1 Introduction

In recent years, the number of landscape fires, the acreage burned by both wildfires and prescribed burning, and the length of the "fire season" in the United States have all substantially increased. Collectively, fires produce about one-third of the total particulate pollution in the United States (U.S.) [4]. However, information about larger wildfires is frequently missing and the contribution from small-scale localized fires is often missing in regional wildfire and smoke inventories. Understanding the effects of fires on the Earth's climate system, including small-scale localized fires, and helping to mitigate the negative effects for society and the environment requires routine, global monitoring, which is only possible from satellites, as ground-based reporting of fires is incomplete and can be missing entirely in certain areas.

The detection, segmentation and tracking of fire and smoke plumes at high spatial resolution are critical for better understanding of the complicated fire-fuel-atmosphere interactions and the impact of fires on air quality and climate. However, the conventional process of manually segmenting images to identify smoke plumes is time-consuming and can be subject to considerable uncertainty. On the other hand, machine learning (ML) tools such as unsupervised and self-supervised approaches might be more well-suited to handle large amounts of data and also can perform label-free image segmentation. The fact that the human-in-the-loop steps of context application and validation occur after the images have been segmented allows for human oversight, while mitigating the need for the uncertainty-laden and extremely labor-intensive act of pixel-by-pixel manual segmentation for tens of thousands of images.

This segmentation pipeline alone works well for segmenting out wildfires and smoke plumes in many differing scenes from various remote sensors. However, it would not be operationally inefficient to run this over every single pixel of remotely sensed data from all sensors used for monitoring wildfires in a region. Instead, we can use this tool in conjunction with others - first as a way to create labels for validation and/or training, and second as a piece of the pipeline to be used for segmentation when the anomalies in question (here wildfires and smoke plumes) have been identified in a subset of the region being monitored. In this way, we can minimize the compute needed, allow for early detection of the ground or even with anomaly detection onboard remote sensing platforms, and run segmentation only on areas that have been pre-screened. In our anomaly detection process, we evaluate conventional ML-based solutions as well as the emerging self-supervised approaches.

2 Overview of the analytic tools for wildfire detection

Efficient detection of anomalies in Earth science processes such as wildfires is exacerbated, first, by the limited or non-existing records of labeled ground truth and, second, by a sophisticated spatio-temporal dependence structure among entities of the underlying process. To address the first challenge, that is, limited or non-existent records of labeled anomalies, anomaly detection associated with wildfires is often viewed as an unsupervised problem. Among such notable DL approaches for anomaly detection are Autoencoding (AE) [2] based on the idea of reconstruction errors; Deep Autoencoding Gaussian Model (DAGMM) [26] which expands AE with the Gaussian Mixture Model, and Variational Autoencoders (VAE) [13] with regularized encoding's distribution. However, such DL methods are restricted in their ability to learn multiple types of interactions among system entities in dynamic settings, e.g., georeferenced sensors. Finally, in the last couple of years, there has been a spike of interest in bringing the power of graph neural networks (GNNs) to anomaly detection tasks on spatio-temporal data [20]. Such GNN-based approaches have the potential to systematically assess complex spatio-temporal interdependencies of the underlying system, but may require additional constraints in terms of data record availability. In this paper, we explore the utility of the broad range of such emerging DL techniques for wildfire detection through the lens of self-supervised learning. Such analysis has the potential to open the door for more sound and reliable usage of ML tools in wildfire risk analytics [15]. With the power of these techniques, we can both get fast and accurate detection of regions with wildfire or smoke inundation and downstream segmentation for further analysis, including, inputs to forecasting models.

3 Overview of Wildfire Data

Satellite Imagery Motivated by the emerging wildfire crisis, we leverage the surface reflectance and top of atmosphere (TOA) brightness data from the National Aeronautics and Space Administration (NASA)/United States Geological Survey (USGS) Landsat 8 and 9 satellites and the Multi-Spectral Instruments over the contiguous United States, along with a corresponding set of labels indicating whether a wildfire is present within a pixel or not. The Harmonized Landsat Sentinel-2 (HLS)[1] project provides surface reflectance and top of atmosphere brightness data from the Operational Land Imagers (OLIs) and the Thermal Infrared Sensors (TIRS) onboard the joint NASA/USGS Landsat 8 and 9 satellites and the Multi-Spectral Instruments (MSIs) on board the ESA's Sentinel-2A and Sentinel-2B satellites. The harmonized datasets provide global data with a repeat cycle of 2-3 days. These data are tiled within the Military Grid Reference System (MGRS) and provided in Cloud Optimized GeoTiff (COG) format. Given the scarce wildfire presence and complex spatio-temporal dependencies in surface reflectance, the datasets are particularly well-suited for benchmarking self-supervised learning tasks both on images and non-Euclidean objects. Our experiments on these benchmark datasets indicate that unsupervised contrastive learning algorithms can capture the structures of views and scenes, and map pixel space of multi-sensor imagery to a high-level embedding space for further downstream tasks, thereby allowing for more cohesive integration of the state-of-the-art ML approaches to wildfire risk analytics.

Here we show an example case using HLS data, but the methodologies described below are instrument agnostic and have been evaluated on multiple orbital and suborbital instruments, allowing for the creation of a sensor web of past, present, and future instruments, and a dataset with increased spatiotemporal resolution.

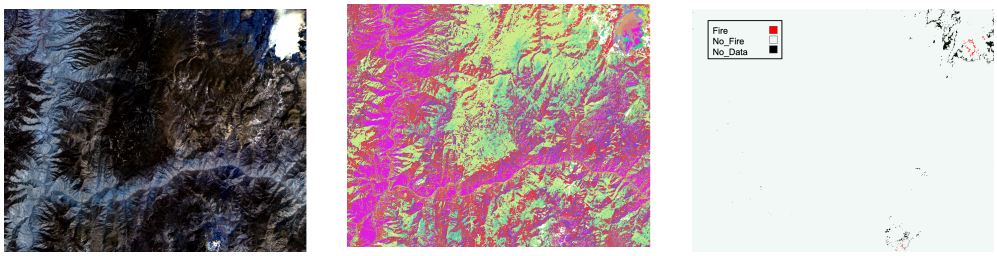
Wildfire Labels Satellite-based instruments have provided comprehensive observations of wildfires and aerosol plumes from wildfires. Most of these instruments do not provide an operational per-pixel wildfire identification dataset, but for those that do, the detection of wildfire and smoke using spaceborne imagery has long required the development of instrument/dataset-specific retrieval algorithms. Such development is labor intensive and requires domain-specific parameters, instrument-specific calibration metrics, and manual effort to track retrieved objects across multiple scenes [8, 25]. Also, many pre-existing operational smoke and fire detection capabilities suffer from issues like the confusion between smoke and clouds and the identification of non-fire objects that are bright or hot [24]. Because of this, we use JPL's Segmentation, Instance Tracking, and data Fusion Using multi-SEnsor imagery (SIT-FUSE), to generate a set of labels – previously validated on numerous pre-existing label sets for land cover types, aerosol types, wildfires, algal blooms, and more [17, 16]. Previous work on the general problem of unsupervised image segmentation appears to have had success in separating the foreground from the background [12, 5], or

[1] https://lpdaac.usgs.gov/products/hlsl30v002/

have only used single bands of input from one type of instrumentation, which is effective for their applications, but does not cover the breadth required here [1]. Other works have aimed to perform tasks, such as outlining buildings and roadways [23], which is not the goal here. Even with more recent breakthroughs in semi-supervised semantic segmentation, like the Segment Anything Model (SAM), a problem-dependent amount of labels is required, and SAM is largely unproven in complex domains like remote sensing [14]. The identification of the necessary size of label sets, generation of per-pixel label sets, and testing of the feasibility of new techniques in more complex domains are all problem-specific and time consuming tasks that can be skipped, given our proposed solution. The lack of need for large new label sets mitigates the costly, labor-intensive work of manually segmenting each pixel within a dataset used for ground truth, a process which is itself error prone, and other previously mentioned supervised learning-related precursors model training. Also, leveraging pre-existing operational products to use as labels for supervised learning tasks will inherently cause them to either lack training set diversity or suffer from issues inerrant to those datasets, like those mentioned above. On the other hand, our proposed ML-based pipeline is well suited to handle large amounts of data, because our unsupervised and self-supervised models can perform label-free image segmentation. The fact that the human-in-the-loop steps of context application and validation occurs after the images have been segmented allowing for human oversight, while mitigating the need for the extremely labor-intensive act of pixel-by-pixel manual segmentation for tens of thousands of images. SIT-FUSE utilizes an unsupervised and self-supervised ML framework, namely Restricted Boltzmann Machines [18, 21] and self-supervised Deep Clustering [11], and allows users to segment instances of objects like wildfires and smoke plumes in single and fused multi-sensor scenes from orbital and suborbital remote sensors with minimal human intervention, in low and no-label environments. The innovative multi-sensor segmentation capabilities allow for the ability to precisely detect anomalous observations from instruments with varying spatial and spectral resolutions, creating a sensor web of pre-existing and future instruments, and has the ability to incorporate model-based data, where appropriate. In our study, each wildfire label set is generated at the input data's native resolution and contains 3 values: -1 impleies no_input_data, 0 implies no_fire, and 1 means $fire$. Figure 3 depicts an example Landsat scene, the associated SIT-FUSE segmentation, and the labels generated from applying context to a subset of the clusters.

4 The Self-Supervised Learning Pipeline

As in the case of most existing remote sensing data products, wildfire presence tends to be scarce and ground truth labels of wildfire occurrence are not readily available, while the labeling process is resource-consuming. This problem is particularly well suited for self-supervised learning tasks. Furthermore, surface reflectance data are characterized by sophisticated nonseparable nonstationary



LandSat RGB reference. SIT-FUSE scene segmentation. Fire labels.

Figure 1: Example input scene (*left*), intermediate segmentation generated by SIT-FUSE (*middle*), and associated fire labels (*right*). The fire labels are created by applying context to a subset of the clusters generated by SIT-FUSE.

spatio-temporal dependencies (i.e., the dependencies over space vary over time and vice versa), making the application of more traditional ML and statistical models relying on Euclidean distances less feasible. As a result, in this section, we focus on illustrating the utility of these approaches as a benchmarking platform for self-supervised learning on images.

We first introduce the background of self-supervised learning and additionally demonstrate that the contrastive learning framework can be practically applied to wildfire detection tasks. A pivotal factor in the success of self-supervised learning is its ability to harness the vast volumes of unlabeled data emerging in the big data era. It signals a shift for deep learning algorithms, moving away from human oversight and towards data-driven self-supervision. That is, self-supervised learning capitalizes on the innate relationships within data as its guiding force, showcasing its adaptability. Here we can summarize the mainstream self-supervision into three general categories: (i) Generative, i.e., training an encoder to encode input x into an explicit vector z and a decoder to reconstruct x from z, (ii) Contrastive, i.e., training an encoder to encode input x into an explicit vector z to measure similarity (e.g., mutual information maximization, instance discrimination), and (iii) Generative-Contrastive, i.e., training an encoder-decoder to generate fake samples and a discriminator to distinguish them from real samples.

In our study, we have focused on contrastive learning. The fundamental concept behind contrastive self-supervised learning is to maximize the mutual information between the representations of two random variables v_1 and v_2 with the joint distribution $p(v_1, v_2)$, i.e., $\max_{f_1,f_2} (\boldsymbol{I}(z_1, z_2))$, where $\boldsymbol{I}(\cdot, \cdot)$ is a sample-based estimator for the accurate mutual information, $z_i = f_i(v_i)$ $(i = \{1, 2\})$ are also random variables and f_i denotes an encoding function (i.e., representation function). In practice, v_1 and v_2 are augmented views of the same data example x (where data augmentation operators include rotating, crop, Gaussian noise, Gaussian blur, color distort, etc) and common practice to alternatively maximize $\boldsymbol{I}$'s lower bound is using a Noise Contrastive Estimation (NCE) [9]. That is, we train networks to amplify the agreement between various views of the same

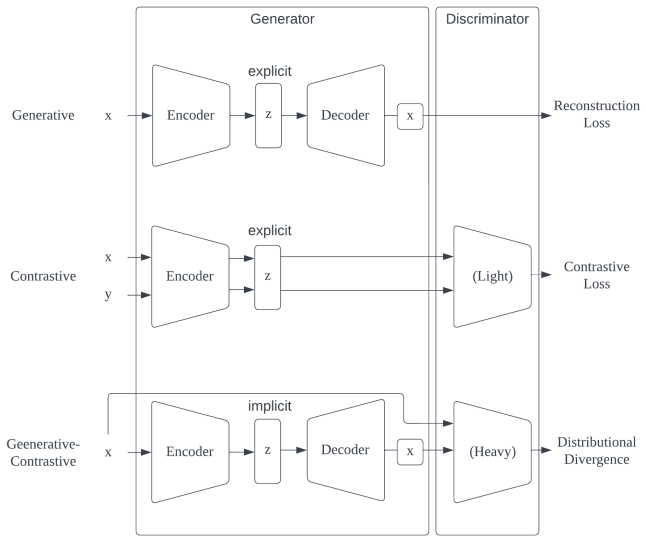


Figure 2: Conceptual comparison between Generative, Contrastive, and Generative-Contrastive methods. [19]

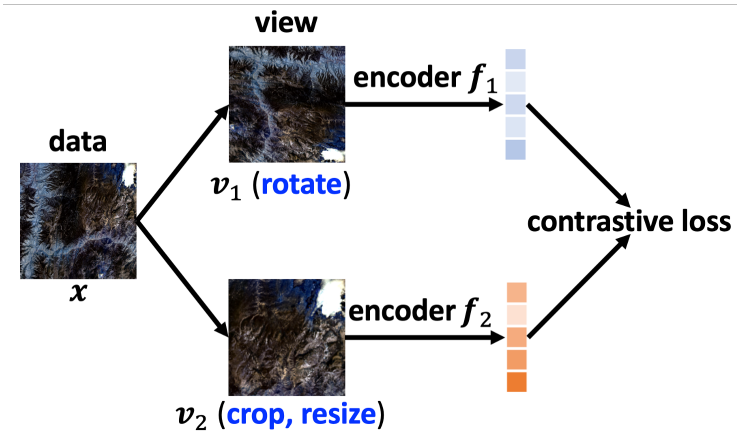


Figure 3: Internal mechanism of contrastive learning: the views provide supervision information to each other.

scene while diminishing the agreement between views from distinct scenes.

5 Experimental results

Two types of widely-used unsupervised ML/DL methods were benchmarked: (i) conventional ML algorithms: (a) KMeans and (b) Random Forest (RF) [3], (ii) CNN-based unsupervised learning models: (a) residual networks (ResNet) [10] (including ResNet-18 and ResNet-50; note that here we fine-tune the pre-trained ResNet models), (b) simple framework for contrastive learning of visual representations (SimCLR) [6] which contains four parts, i.e., data enhancement module, encoder module, and projection head and task-specific contrast loss function, and (c) vision transformer (ViT) [7].

5.1 Validation via Reconstruction Errors

Anomaly detection is a classification problem to identify abnormal patterns. An intuitive approach is to model the distribution of regular patterns, where data that does not agree with the distribution of regular patterns can be identified as irregular. However, in the absence of labels distinguishing normal data from abnormal data, applying extensive domain expertise and professional knowledge becomes necessary. Therefore, an unsupervised novelty detection method is often the optimal choice in such instances. Inspired by the success of the autoencoders [22], here we use the above benchmarking models to extract latent feature representations and utilize the reconstruction error of autoencoder as a measurement of anomaly. As such, we apply the autoencoder framework over a latent feature representation to model the normal distribution, and irregular patterns will be distinguished by large reconstruction errors. More specifically, to learn the representations or reconstructions of its original input as close as possible, we utilize the sparse autoencoder (i.e., a single-layer autoencoder). The sparse autoencoder consists of an encoder, a hidden layer, and a decoder, which tries to find common data representation from the input. For an input sample $\boldsymbol{z} \in \mathbb{R}^d$, the encoding function $f(\boldsymbol{z})$ can be expressed as $f(\boldsymbol{z}_i) = \sigma(\boldsymbol{W}^{(e)}\boldsymbol{z}_i + \boldsymbol{b}^{(e)})$, where $\boldsymbol{W}^{(e)}$ and $\boldsymbol{b}^{(e)}$ represent the trainable weight matrix and the bias vector respectively, and $\sigma(\cdot)$ is the non-linear activation function applied component-wise. Then, the hidden representation is mapped via a decoding function to reconstruct the original input as $\tilde{\boldsymbol{z}}_i = \sigma(\boldsymbol{W}^{(d)} f(\boldsymbol{z}_i) + \boldsymbol{b}^{(d)})$, where $\tilde{\boldsymbol{z}}$ is the vector of output values, and $\boldsymbol{W}^{(d)}$ and $\boldsymbol{b}^{(d)}$ are the trainable weight matrix and the bias vector of the decoder respectively. Thus, given a latent embedding $\boldsymbol{Z} \in \mathbb{R}^{N \times d}$, the sparse autoencoder iteratively updates the representation parameters by minimizing the reconstruction error $\sum_{i=1}^{N}(||\boldsymbol{z}_i - \tilde{\boldsymbol{z}}_i||)^2$.

Table 1: Performance comparison for wildfire detection in terms of the reconstruction error (RE) between SOTA image representation learning methods and ML algorithms.

Model	KMeans	RF	ResNet18	ResNet50	SimCLR	ViT
Recon.Error	0.1238 ± 0.0001	0.1233 ± 0.0001	0.0156 ± 0.0001	0.0230 ± 0.0003	0.0001 ± 0.0001	0.0042 ± 0.0053

Table 1 shows the comparison in performance among 6 models in terms of the reconstruction error over the SIT-FUSE dataset. As expected, neural network-based models (i.e., ResNet18, ResNet50, SimCLR, and ViT) always substantially outperform more traditional ML methods (i.e., KMeans and RF), with a few orders of magnitude.

5.2 Validation against Annotated Ground Truth Wildfire Labels

Tables 2 displays comparative performance among 6 models over the ground truth labels generated by SIT-FUSE. Similar to the case of evaluation against reconstruction errors, as might be expected, neural network-based models significantly outperform ML methods, both in terms of accuracy and variability. Interestingly, while the accuracy of SimCLR is slightly lower, i.e., 90.40%, its variability is substantially lower than that of all models (i.e., 0.01). We hypothesize that these phenomena may be due to the following reasons: (i) from a classification perspective, in our experiments, superpixels can produce promising quality segmentations by using graph cuts, i.e., capturing diverse visual patterns within an image, (ii) pre-defined superpixels may lack the control of quality and compactness of segments which increases uncertainty especially for imbalanced learning. In contrast, ViT exhibits comparable performance to other neural network models, achieving an accuracy of 90.46%. However, its variability notably surpasses that of all models, standing at 0.56. We hypothesize that this phenomenon may be attributed to ViT's analytical approach. Unlike traditional convolutional neural network (CNN) methods, transformers, such as ViT, emphasize the analysis of global information in the dataset. When detecting fire pixels in satellite images, focusing on specific subsets where the designated fire region is prominent can stabilize the model, providing a plausible explanation for ViT's higher variability compared to its counterparts.

Table 2: Accuracy (%) of SOTA image representation learning methods and ML algorithms.

Model	**KMeans**	**RF**	**ResNet18**	**ResNet50**	**SimCLR**	**ViT**
Accuracy	85.25±0.19	85.20±0.26	90.50±0.49	90.52±0.27	90.40±0.01	90.46±0.56

Our results demonstrate the utility of contrastive learning for wildfire region identification, thereby suggesting that the early detection of wildland fires could be greatly improved both in terms of speed and efficacy. Furthermore, even in the absence of the ground truth labels, given the relatively low reconstruction errors shown in Tables 1, we can conclude that the rich information content of the high-resolution multispectral imagery tends to be captured well by the CNN models. These important findings open the door for exploration of the encoded feature spaces and model trunks for further ML use cases and interpretability studies of Earth Science processes.

Limitations Although remotely sensed imagery and the SIT-FUSE generated wildfire labels offer valuable insights into the distribution and variability of wildfires, they are not without limitations. Firstly, while the approach is not limited to a specific geographic are, in this experiment the dataset's spatial domain is limited to the contiguous United States. Further experimentation and label generation will need to be done to evaluate the performance of these techniques on broader swaths of the globe. Secondly, this dataset and experimentation primarily focuses on active fire detection, despite the importance of utilizing satellite imagery for studying pre- and post-fires. Additionally, surface reflectance and TOA brightness from satellites contain considerable variability due to sensor characteristics, leading to potential uncertainties in the wildfire label. Adding uncertainties alongside each segmentation is an ongoing research task for the SIT-FUSE team. Despite these limitations, these datasets and the fire detection capabilities from them remain a valuable tools for fire monitoring, but complementary data sources and the application of enhanced ML models are necessary to overcome these limitations.

6 Conclusions and discussion of future work

This chapter has discussed the challenges associated with effectively detecting anomalies in wildfires and has outlined the proposed solutions in the form of the emerging concepts of self-supervised learning. In particular, the two primary challenges for wildfire detection are the scarcity of labeled data and the intricate spatio-temporal dependencies within these events. Recently, there has been a burgeoning interest in harnessing GNNs for spatio-temporal anomaly detection. In light of this, the concept of self-supervised learning with its potential to enhance wildfire detection by swiftly and accurately identifying wildfires, smoke inundation, and supporting further analysis and forecasting in wildfire risk analytics, offers a novel framework to mitigate the fundamental challenges on the way of more accurate wildfire detection in the absence of reliable ground truth data and under the scenarios of complex spatio-temporal dependencies.

The chapter has also explored the utilization of satellite data, including information from NASA/USGS Landsat 8 and 9 satellites, and ESA's Sentinel-2A and Sentinel-2B satellites, for wildfire detection and labeling. These datasets are particularly well-suited for self-supervised learning and benchmarking on images and non-Euclidean objects. To address the challenge of labeling wildfire data from satellite imagery, the SIT-FUSE system employs unsupervised and self-supervised ML techniques to generate labels for wildfires and smoke plumes. This approach is adept at handling extensive datasets, mitigating the need for manual pixel-by-pixel labeling. The result is a precise wildfire detection method that provides labels indicating whether a region contains no data, no fire, or a wildfire at native resolution.

In conclusion, we hope that the presented findings will open the door to quicker and more efficient wildfire identification using ML models, thereby enhancing wildfire risk analytics.

Acknowledgments

This work has been supported in part by the NASA AIST grant 21-AIST21_2-0059. In addition, Y.C., H.L., and N.L. have been supported by the NSF Grant DMS-2335846/2335847 and Y.C. has also got support from the NSF grant TIP-2333703. Part of this material is also based upon work supported by (while serving at) the NSF. The views expressed in the article do not necessarily represent the views of NSF .

References

[1] Iman Aganj, Mukesh G Harisinghani, Ralph Weissleder, and Bruce Fischl. Unsupervised medical image segmentation based on the local center of mass. *Scientific Reports*, 8(1):13012, 2018.

[2] Charu C. Aggarwal and Saket Sathe. Theoretical foundations and algorithms for outlier ensembles. *SIGKDD Explor. Newsl.*, 17(1):24–47, sep 2015.

[3] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[4] Marshall Burke, Anne Driscoll, Sam Heft-Neal, Jiani Xue, Jennifer Burney, and Michael Wara. The changing risk and burden of wildfire in the United States. *Proceedings of the National Academy of Sciences*, 118(2):e2011048118, 2021.

[5] Mickaël Chen, Thierry Artières, and Ludovic Denoyer. Unsupervised object segmentation by redrawing. In *Proceedings of the Advances in Neural Information Processing Systems*, volume 32, 2019.

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning*, pages 1597–1607. PMLR, 2020.

[7] 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 *roceedings of the International Conference on Learning Representations, year=2021, url=https://openreview.net/forum?id=YicbFdNTTy.*

[8] L. Giglio and C. Justice. MOD14A2 MODIS/Terra Thermal Anomalies/Fire 8-day L3 Global 1km SIN Grid V006. *https://doi.org/10.5067/MODIS/MOD14A2.006*, 2015.

[9] Michael Gutmann and Aapo Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and*

statistics, pages 297–304. JMLR Workshop and Conference Proceedings, 2010.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[11] Xu Ji, João F Henriques, and Andrea Vedaldi. Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9865–9874, 2019.

[12] Asako Kanezaki. Unsupervised image segmentation by backpropagation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1543–1547. IEEE, 2018.

[13] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv e-prints*, page arXiv:1312.6114, December 2013.

[14] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv:2304.02643*, 2023.

[15] Lee H. Garay M. Guo Z. Dixon M. Chen Y. Segovia Dominguez I. Gel Y.R. LaHaye, N. Automatic smoke plume and wildfire instance tracking across multi-sensor scenes. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2023.

[16] N. LaHaye, J. Ott, M. J. Garay, H. M. El-Askary, and E. Linstead. Multimodal object tracking and image fusion with unsupervised deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8):3056–3066, 2019.

[17] Nicholas LaHaye, Michael J. Garay, Brian D. Bue, Hesham El-Askary, and Erik Linstead. A quantitative validation of multi-modal image fusion and segmentation for object detection and tracking. *Remote Sensing*, 13(12), 2021.

[18] Renjie Liao, Simon Kornblith, Mengye Ren, David J. Fleet, and Geoffrey Hinton. Gaussian-Bernoulli RBMs without tears. *arXiv:2210.10318*, 2022.

[19] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge amp; Data Engineering*, 35(01):857–876, jan 2023.

[20] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z. Sheng, Hui Xiong, and Leman Akoglu. A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2021.

[21] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the International Conference on Machine Learning*, 2010.

[22] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[23] Anderson Reis Soares, Thales Sehn Körting, Leila Maria Garcia Fonseca, and Alana Kasahara Neves. An unsupervised segmentation method for remote sensing imagery based on conditional random fields. In *Proceedings of the IEEE Latin American GRSS & ISPRS Remote Sensing Conference*, pages 1–5, 2020.

[24] Aaron van Donkelaar, Randall V. Martin, Robert C. Levy, Arlindo M. da Silva, Michal Krzyzanowski, Natalia E. Chubarova, Eugenia Semutnikova, and Aaron J. Cohen. Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the moscow fires in 2010. *Atmospheric Environment*, 45(34):6225–6232, 2011.

[25] Jing Wei, Yiran Peng, Jianping Guo, and Lin Sun. Performance of modis collection 6.1 level 3 aerosol products in spatial-temporal variations over land. *Atmospheric Environment*, 206:30–44, 2019.

[26] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *Proceedings of the International Conference on Learning Representations*, 2018.