IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10281964

AN ANALYSIS OF THE GAP BETWEEN HYBRID AND REAL DATA FOR VOLCANIC DEFORMATION DETECTION

Teo Beker[1,2], Qian Song[1], Xiao Xiang Zhu[1]

[1] Data Science in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany
[2] Remote Sensing Technology Institute (IMF), Weßling, German Aerospace Center (DLR), Germany

ABSTRACT

Recently deep learning models were applied to detect fast short-term volcanic deformations using interferometric synthetic aperture radar (InSAR) data. However, volcanic deformation detection is limited by the availability of real positive samples. In previous work, we used hybrid synthetic-real InSAR deformation maps set to train an InceptionResNet v2 model capable of detecting deformations down to 5 mm/year in real set. However, our model also reported false positive detections. One possible reason is the data distribution gap between the real and hybrid sets. In this paper, an experiment is conducted to analyze the gap between the hybrid and real sets that resulted in false positives. Three subsets of the fine-tuning set are created based on t-SNE analysis using different sampling strategies. The classification model is fine-tuned using these subsets. The results show that the strategy of removing only the most confusing examples and keeping the larger data set size reduces the false positive rate from 32.29% to 27.01%.

Index Terms— Volcano Deformation Detection, t-SNE, XAI, Data Gap

1. INTRODUCTION

One thousand five hundred volcanoes are considered active, and 800 million people live less than 100km from them. Therefore, forecasting volcanic activity and eruptions are of great significance. While in developed countries and more densely populated areas, there are often seismic or global navigation satellite system monitoring stations, most volcanoes do not have on-site monitoring. Past decades saw synthetic aperture radar (SAR) used for monitoring volcanic deformation from space. Initially, expertise and significant time were needed to produce and analyze the data, but today systems relying on artificial intelligence (AI) for automatically detecting volcanic deformations are becoming more common.

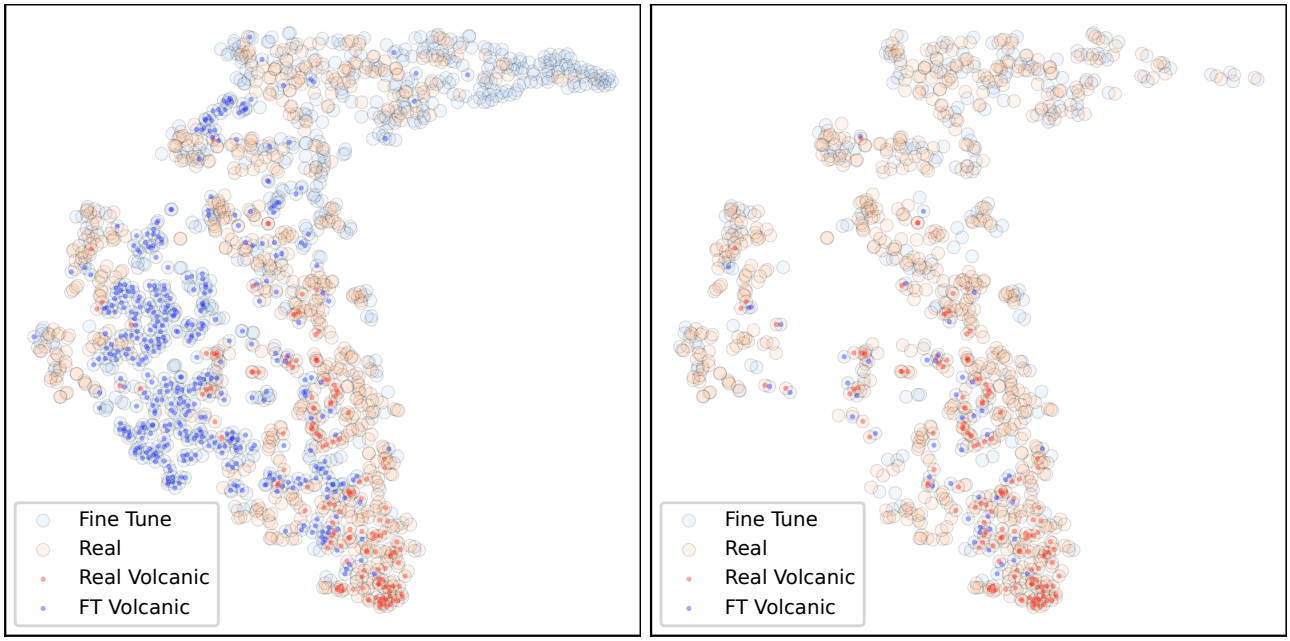
Most AI systems focused on short-term intense volcanic deformations [1, 2], where single interferograms can be utilized, which are more commonly present. Few works focused on persistent scatterer interferometric aperture radar (PSInSAR), which allows the detection of very subtle volcanic deformations but needs significantly more data processing. The commonly used result of this approach is cumulative deformation maps representing the deformations of a whole observed period in a single image, significantly reducing the data that can be used for training AI models.

Different AI models are generated to detect volcanic deformations in interferometric SAR (InSAR) data. A common approach is using classification models, but most of these approaches face similar data-related problems. Despite the satellite revisit period becoming shorter, volcanic deformations are not a very common phenomenon, meaning issues forming the training sets. In the past, this was solved using synthetic data, large quantities of unlabeled data, fine-tuning with hybrid synthetic-real data [3, 4], or image augmentations. However, many false positives were detected (32.29% in [4]).

The main issues are small quantities, underrepresented data diversity and sparse distribution [5], and confusing and unrealistic samples in the training set. Previous studies show these problems are further exaggerated when using PSInSAR data [3, 4]. We use PSInSAR data from South America's Central Volcanic Zone (CTZ) from 2015 to 2020 and InceptionResNet v2 model to detect volcanic deformations down to 5mm/year [6]. We solved the limited training data, using synthetic data for training and a small hybrid real-background and synthetic volcanic deformations data for fine-tuning the model. The model is tested on unseen real data covering 46 volcanoes in the observed region. While the model detects all volcanic deformations, it suffers from an increased false positive rate. This paper explores the gap between the hybrid fine-tuning set and the real data and attempts to narrow it. For this, we generate three subsets of initial fine-tuning set using different sample selection strategies and compare the achieved results.

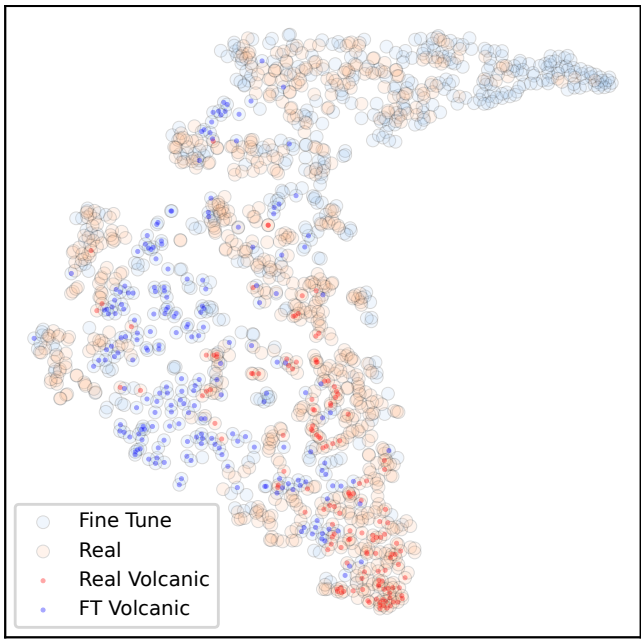
2. METHODOLOGY

The research area is the central volcanic zone of the Andes from 2014 to 2020. The data is processed into high-accuracy deformation maps [7, 8], which contain the deformation over

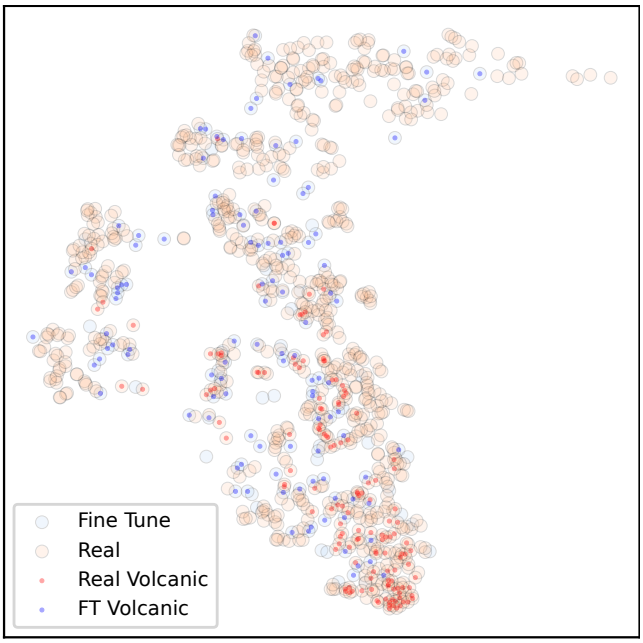


(a) Original fine-tuning set

(b) S1) The closest sample to a real sample



(c) S2) More similar to a real sample of same label



(d) S3) More similar to a sample of same label

Fig. 1: T-SNE projection of the fine-tune subsets and real test set features. (a) Samples of the whole fine-tuning set (FT) and real test set projected in t-SNE feature reduced space. (b-d) subsets of the FT selected by the sample similarity in feature-reduced space. In (d), an eliminated samples subset is shown.

the whole period in each image. Thus, the data is rich in deformation signals such as slope processes, salt lakes, and volcanic deformations. Besides the deformation signal, residual tropospheric noise is present, as it could not be entirely corrected, which constitutes additional challenges. The data surrounding 46 known volcanoes are kept for the real test set, and the rest is patched to create a hybrid set using synthetic volcanic deformations to make a positive class.

The InceptionResNet v2 classification model is trained on around 300,000 synthetic data samples and fine-tuned by 836 hybrid synthetic volcanic deformations samples and real background data samples. While all the volcanic deformations are detected, the false positive rate (FPR) is high. This could result from sparse distributions of the fine-tuning set and confusing samples [5]. Furthermore, hybrid synthetic-real data implies some artifacts and confusing examples. Since the fine-tuning set is small, these samples can significantly impact the final model. We propose improving the models through better fine-tuning sets, and we test three strategies for generating fine-tuning sets. To measure the similarity between the samples, we use Euclidean distance in a 2D space embedded by t-SNE, which will be detailed in section 2.1.

2.1. Subsets strategies

T-SNE [9] is a dimensionality reduction algorithm. We use it to visualize and compare the features of fine-tuning sets and real samples. First, the features are extracted from the last convolutional layer of the model, amounting to 1536 features for each sample. Then, all samples from the sets are used with t-SNE to reduce the feature space to two dimensions.

To improve the quality of the fine-tuning set, we select three subsets from the hybrid samples to fine-tune the network [10], and compare their performances with the original model. Using Euclidean distances between t-SNE embedded features as a similarity measure, three subsets are created: S1 uses only the most similar samples to each of the real samples, S2 removes all the samples which are more similar to the real samples of an opposite class than to the correct class, and S3 removes only the samples most similar to the opposite class (Figure 1). To keep the class balance, each subset has been used as is or undersampled and oversampled [11].

2.2. Fine-tuning

The model is improved by freezing the whole model and fine-tuning the last or last two layers. Each model's classification threshold is set to detect 94.37% of positive samples (TPR) for equal comparison. To compare the models, we focus on the area under the curve receiver operating characteristic (AUC ROC) and FPR while setting the common true positive rate (TPR). To compare the data, we use the features extracted from the last convolutional layer of the model and transform it into two dimensions using t-SNE for visualization.

3. RESULTS

Table 1 shows the metrics achieved by models fine-tuned using the subsets. Fine-tuning the last two layers extended the models' capacity to learn the fine-tuning set, often increasing the validation accuracy but not the models' performance on the real test set, implying the overfitting due to the small size of the sets.

As expected, the hybrid positive samples constitute most of the confusing samples. There were significantly fewer confusing backgrounds than volcanic deformation samples in every tested scenario.

The results of fine-tuning using S1 and S2 generally show a decline in performance, probably because of a noticeable reduction in set size (8% to 88%) and diversity (37% to 88%). An exception is the oversampled S1 set, which achieved 88.04% AUC ROC, an increase from the original 86.44%. In addition, it contained the highest quality samples, using only 27.63% of the original data and eliminating all the confusing examples. But this model underperformed at the set threshold, having a higher FPR.

The best performance was achieved by fine-tuning the last layer of the model using subset S3 because of eliminating the most confusing samples while maintaining a large enough quantity of samples. While using the original subset S3 gives the most significant area under the curve receiver operating characteristic (AUC ROC), applying the classification threshold, the undersampled subset S3 produces a marginally better model. Even though the subsets are 17.58% and 27.75% smaller than the original set, they reduce the false positive rate (FPR) from 32.39% to 27.35% and 27.01%, respectively.

4. CONCLUSION

This paper uses t-SNE to modify the fine-tuning sets and extract additional model performance. T-SNE is a valuable tool for identifying the similarity between the samples. Therefore, using it to eliminate the most confusing samples significantly boosts the model performance. On the other hand, removing too many samples leads to overfitting and a decline in model performance.

Three strategies for reducing the fine-tuning set and increasing its quality are tested. It is shown that fine-tuning sets often can be reduced even with an increase in the model's performance. The most consistently helpful strategy for increasing model performance is eliminating confusing examples and maintaining the larger sample count. For example, the undersampled S3 strategy produced a 27.75% smaller fine-tuning set and a model with a 5.38 % smaller FPR. In summary, the experiments uncovered the gap between the real and fine-tuning sets, which affected the model's performance. In the future, efforts will be paid to increase the training data.

Table 1: Detection metrics of InceptionResNet v2 with different fine-tuning set subsets. S1 contains only the most similar samples to each real sample (P 50/N 181). In S2, all the confusing samples are removed (P 146/N 386), and in S3, only the samples most similar to the opposite class are removed (P 302/N 387).

Model	TP	FP	FN	TPR ↑	FPR ↓	Accuracy ↑	Precision ↑	F1 ↑	AUC ROC ↑
Original	134	205	8	94.37	32.39	72.52	39.53	55.71	86.44
Fine-tuning the last layer									
S1	134	288	8	94.37	45.50	61.81	31.75	47.52	82.75
S1 oversampled	134	242	8	94.37	38.23	67.74	35.64	51.74	88.04
S1 undersampled	134	278	8	94.37	43.92	63.10	32.52	48.38	82.64
S2	134	241	8	94.37	38.07	67.87	35.73	51.84	86.03
S2 oversampled	134	237	8	94.37	37.44	68.39	36.12	52.24	85.74
S2 undersampled	134	259	8	94.37	40.92	65.55	34.10	50.09	87.86
S3	134	173	8	94.37	27.33	76.65	43.65	59.69	**90.23**
S3 oversampled	134	194	8	94.37	30.64	73.94	40.85	57.02	88.87
S3 undersampled	134	**171**	8	94.37	**27.01**	**76.90**	**43.93**	**59.96**	88.83

5. REFERENCES

[1] Nantheera Anantrasirichai, Juliet Biggs, Fabien Albino, and David Bull, "A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets," *Remote Sensing of Environment*, vol. 230, pp. 111179, 2019.

[2] M Gaddes, A Hoopera, and F Albinob, "Simultaneous classification and location of volcanic deformation in sar interferograms using deep learning and the volcnet database," 2021.

[3] Teo Beker, Homa Ansari, Sina Montazeri, and Qian Song, "Detection of volcanic deformations in insar velocity maps-a contribution to tecvolsa project," 2022.

[4] Teo Beker, Homa Ansari, Sina Montazeri, Qian Song, and Xiao Xiang Zhu, "Fine-tuning cnns for decreased sensitivity to non-volcanic deformation velocity signals," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, pp. 85–92, 2022.

[5] Qian Song, Feng Xu, Xiao Xiang Zhu, and Ya-Qiu Jin, "Learning to generate sar images with adversarial autoencoder," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[6] Teo Beker, Homa Ansari, Sina Montazeri, Qian Song, and Xiao Xiang Zhu, "Explainability analysis of cnn in detection of volcanic deformation signal," in *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022, pp. 4851–4854.

[7] David T Sandwell and Evelyn J Price, "Phase gradient approach to stacking interferograms," *Journal of Geophysical Research: Solid Earth*, vol. 103, no. B12, pp. 30183–30204, 1998.

[8] Alessandro Parizzi, Francesco De Zan, Fernando Rodriguez Gonzalez, Homa Ansari, Giorgio Gomba, Ramon Brcic, and Michael Eineder, "Processing and performance of large scale deformation monitoring with insar stacks.," in *Geophysical Research Abstracts*, 2019, vol. 21.

[9] Laurens Van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[10] Peng Peng and Jiugen Wang, "How to fine-tune deep neural networks in few-shot learning?," *arXiv preprint arXiv:2012.00204*, 2020.

[11] Roweida Mohammed, Jumanah Rawashdeh, and Malak Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *2020 11th international conference on information and communication systems (ICICS)*. IEEE, 2020, pp. 243–248.