FINE-TUNING CNNS FOR DECREASED SENSITIVITY TO NON-VOLCANIC DEFORMATION VELOCITY SIGNALS

Teo Beker[1]*, Homa Ansari[1], Sina Montazeri[1], Qian Song[1], Xiao Xiang Zhu[1, 2]

[1] Remote Sensing Technology Institute (IMF), Weßling, German Aerospace Center (DLR), Germany -
(Teo.Beker, Qian.Song, Xiaoxiang.Zhu)@dlr.de, Homa.Ansari@icloud.com, Sina.Montazeri1988@gmail.com
[2] Data Science in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany

KEY WORDS: InSAR, Velocity maps, Volcanic deformations , InceptionResNet v2, Fine-tuning, Synthetic data, Deep learning

ABSTRACT:

Monitoring volcanic deformations allows us to track dynamic states of a volcano and to know where an eruptions could happen. Spaceborne Synthetic Aperture Radar (SAR) and SAR interferometry (InSAR) techniques created an opportunity to track volcanoes globally, even in inaccessible regions without ground measuring stations.This paper proposes a convolutional neural network (CNN) for detection of volcanic deformations in InSAR velocity maps. We had only a small amount of velocity maps over the region of central South American Andes, therefore the synthetic data are used to train the model from scratch. In the region of interest, the velocity maps contain the patterns of salt lakes and slope induced signal which confuse CNN models trained on synthetic data.In order to bridge the gap between the synthetic and real data, the hybrid synthetic-real data set is used for fine-tuning the model. The hybrid set consists of the real background signal data and synthetic volcanic data. Four fine-tuning sets which were created by different combinations of the original hybrid data, the filtered hybrid data, and simulated data have been used and compared with each other. Besides, we compared four fine-tuning approaches to determine where and how to fine-tune the model. Results show significant improvement in performance by majority of the approaches, and training the last or last two layers have given the best results. In addition, using the FT1 (containing only hybrid set), and FT4 (containing all sets) improved the area under the curve receiver operating characteristic (AUC ROC) from 55% to 86% and 88% respectively.

1. INTRODUCTION

The movement of the ground has been tracked for centuries, often motivated by disaster prediction or mitigation of big cataclysmic events like earthquakes, volcanic eruptions or explosions. Today we have sophisticated devices to follow such events on the ground. Since relatively recently, we are able to track such events from space using Synthetic Aperture Radar Interferometry (InSAR) which enables us to track geophysical events happening on the global level, even in regions without ground stations. With this useful tool, time series processing approaches like persistent scatterers interferometry (Ferretti et al., 2000, Adam et al., 2003), allow even to detect millimeter scale deformations with confidence.

With the rising number of spaceborne Synthetic Aperture Radar (SAR) satellites, the data are constantly produced to approach the idea of near-real-time Earth monitoring. To evaluate this large amount of data in a timely manner, an automatic way of detecting the signals is essential. Deep learning (DL) models, which have been widely used for target detection, could be used for checking the data and flagging it as deforming, or in extreme cases sending early warnings. To this end, DL models need to be trained on a significant data set that is large both in number and variance. Since there is not much recorded data of rare geophysical events like volcanic eruptions or earthquakes, such a reliable and accurate automated model is challenging to achieve.

In the field of DL for SAR volcanic deformations detection, research is focusing on the detection of short and long-term deformations. In (Anantrasirichai et al., 2018) it is demonstrated that short-term large-scale volcanic deformations can be detected using Alexnet architecture. Further, in (Anantrasirichai et al., 2019b) the authors show that additional performance can be extracted by using simulations for training the model. (Valade et al., 2019) developed a complete multi-sensor system for tracking the volcanic activity, among which Sentinel-1 is used for deformation detection using interferograms and a CNNs. The architecture based on VGG16 was developed to perform classification and detection tasks at the same time in (Gaddes et al., 2021). The multiple combinations of input data were also compared and wrapped data in all three channels gave the best performance. The self-supervised techniques have been utilized in (Bountos et al., 2021) in order to improve performance and robustness of the classification model.

The long term deformation detection using the time series generally focused on filtering the transient noise, and increasing the signal-to-noise ratio (SNR). In (Gaddes et al., 2018), it is done by separating independent components using Independent Component Analysis (ICA) algorithm, and in (Gaddes et al., 2019) the results are improved by developing the ICASAR algorithm . An overwrapping approach has been tested in (Anantrasirichai et al., 2019a), in order to improve the sensitivity of the model and to detect smaller deformations. There has also been experiments using simulation data in (Sun et al., 2020) to train an encoder-decoder architecture to tackle this task.

To the best of our knowledge, we are the first to utilize velocity maps for long-term automatic detection of volcanic deformations. Velocity maps are created using a large interferometric stack and processing them using persistent scatterers (Crosetto et al., 2016) or distributed scatterers (Even and Schulz, 2018), thus preserving high temporal stability and extracting the subtle motions throughout the time period the data were obtained. The product of these approaches is a velocity map, which shows

* Corresponding author

the deformations in time [cm/year] during observation, able to detect even orders of magnitude smaller deformations than interferogram pairs (Parizzi et al., 2019). In this paper, the state of the art method was used to reduce the error to mm level, called Eigendecomposition-based Maximum-likelihood-estimator of Interferometric phase (EMI) (Ansari et al., 2019).

The usage of velocity maps results in a smaller amount of data as well as training samples, which makes the training of a DL model more challenging. In the velocity maps most of the atmospheric phase screen (APS) is removed and the visibility of subtle deformations on the surface is increased. These region and data specific subtle deformations, of salt lakes and slope induced signal, have not been accounted in previous work. These patterns create a gap between the simulations and real data.

In this paper we propose an approach to train a DL model for the detection of subtle volcanic deformations in a specific region of central South American Andes. The goal of the model is to mark the images containing the patterns of volcanic deformations, and it is intended to be used in pipeline, flagging volcanic deformations for later expert in-depth analysis. The main challenge in training such a model arises from extremely limited data, known as few shot learning problem. No real volcanic data can be used for training the models. To solve the problem of little training data, the synthetic data sets of the residual atmospheric noise and volcanic deformations are created. Residual atmospheric noise is simulated according to the Monte Carlo approach (Gaddes et al., 2018), using the values extracted from the modeled variograms of the real data interferograms. The volcanic deformation simulations were created using compound dislocation models (Nikkhoo et al., 2016).

The second encountered challenge arises from the low intensity of volcanic deformations in the region and region-specific patterns observed in long-term velocity maps. Even though the velocity map processing almost eliminates the APS, the salt lakes and slope-induced signal in this region constitute the noise that the model needs to learn to handle. These patterns can confuse the model which does not account for them, with their similarity to volcanoes in patterns and intensity. The region and the specific patterns can be seen in Figure 1. A fine-tuning step using hybrid real-synthetic data are suggested in this work as a solution for bridging the gap between the synthetic training data and real data. Four different approaches of fine-tuning the model and four approaches to creating the fine-tuning set are applied on the model trained on synthetic data, and are tested and compared.

2. DATA

2.1 Real data

The data used to test the models is based on Sentinel-1 SAR imagery obtained from 2014 to 2020 over South American central Andes. The data are corrected for solid Earth tide, tropospheric and ionospheric delay and preprocessed into velocity maps with temporal baseline of 4.5-5.5 years based on about 160 time series images per frame. The velocity maps were transformed to cumulative deformations over the experimental period and wrapped using the Sentinel-1 wavelength: ~ 5.546 cm. The scale of cumulative volcanic deformations is 2-9 cm over the period of about 5 years, or about 0.4 - 1.8 cm/year.

The data are further cut into frames centered on known volcanoes in the region. Frames had a size of 102.4 km by 102.4 km and a resolution of 512 by 512 pixels, effectively giving spatial resolution of 200m per pixel. The data are preselected to not have more than 5% missing values. This resulted in 31 frames centered over 24 different volcanoes and covering ~ 46 volcanoes in the region.

Since 512 pix images are too large to feed to CNN models, the frames are patched into 256 by 256 pix, with step of 64 pixels, meaning that there was 75% overlap between the patches. The overlap is beneficial since the volcanoes can be of large size and they need to be completely contained in a single patch for detection because simulated training data does not account for partial detection. The set created this way totaled 775 images.

This amount of data were limited and had highly imbalanced classes with/without volcanic deformations. The patches with present volcanic deformations are scarce and could not be spared for training. This set is kept for final testing only.

2.2 Simulated data

To train the model, synthetic data were needed. The synthetic data set is created to have equal part samples with volcanic deformations (deformation class) and without volcanic deformations (non deformation class). Residual atmospheric noise is used in both classes as basis, but has been separately simulated for ascending and descending track. Depending on the time when the satellites of orbit arrive to the observed location, the charge in the atmosphere changes (**?**). Over the observed region descending track arrived at time with less charge, giving lower noise ($4.6025 \pm 0.1719mm$ cumulative deformation over 5 years) than ascending tracks ($8.299 \pm 1.9365mm$).

A side experiment showed the models making spurious correlations and thus giving lower performance, if a fraction of the volcanic deformation samples were obscured by atmospheric noise. Therefore this possibility was excluded by making volcanic deformations larger than $\mu + 3\sigma$ of atmospheric noise. 23631 simulations of diverse volumetric source models corresponding to sill, dyke, pipe and other volcanic deformations were created ranging in scale from 1.5 cm for ascending track and 0.6 cm for descending track to 10 cm using compound dislocation models (Nikkhoo et al., 2016). The simulations have spatial extent of 51.2 km by 51.2 km, and spatial resolution of 200m, same as real data. The volcanic deformations were positioned at the center of the images. Because of overlapping patching, position of volcano in the image should not matter, but the model will not be trained for partial detections.

Since real data is corrected for tropospheric and ionsopheric delays, assumption is made that stratified atmosphere (Gaddes et al., 2018) is eliminated, and only residual turbulent atmosphere needs simulation. The generation of residual atmospheric noise was done using Monte Carlo approach from (Gaddes et al., 2018), according to the extracted parameters from $\sim$ 10000 variograms calculated from the collected Sentinel-1 data. The simulated atmospheric noise has been separately generated for ascending and descending track as the noise has significantly different intensity depending on the track. With respect to volcanic deformation simulations the double amount of residual atmospheric noise simulations have been generated. Half to be used in deformation class and half to use for non deformation class.

After the simulation data were created it needed to be projected along appropriate satellites lines of sight. Each set of 47,262

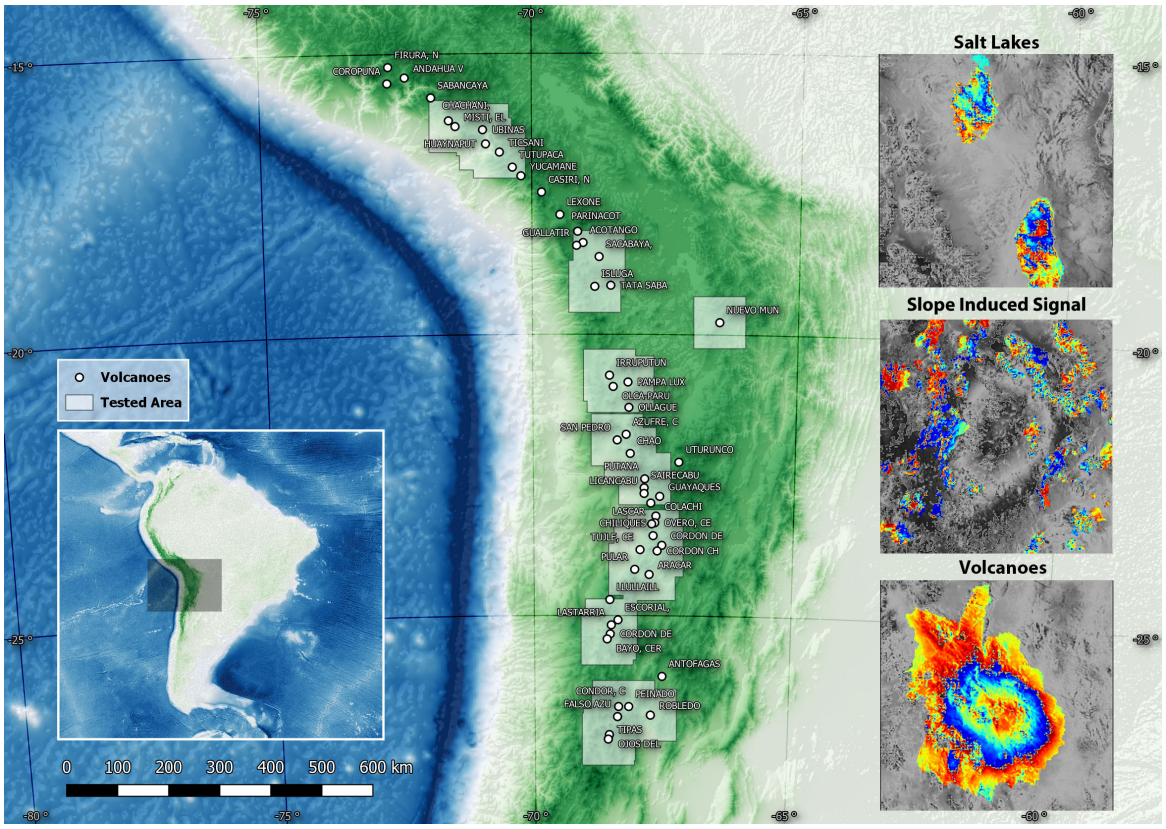

Figure 1. The region of interest in the central South American Andes with marked out locations of volcanoes and positions of the frames used in the real test set. On the right are visible the patterns of salt lakes and slope induced signal which have been encountered in the velocity map data.

samples, half with volcanic deformation and half without has been projected using the one set of the extracted parameters. The parameters were extracted from all of the used real data frames. 7 sets of LOS parameters (east, north, and vertical displacement components) were selected, 2 sets of ascending and 5 of descending tracks. To project the simulated data into these 7 sets of LOS parameters, for each set of parameters an independently simulated set of data were created, totaling 330,834 samples.

This simulated data are used to train the convolutional neural network (CNN) from scratch. The training set is created by random selection of 90% of data, or 297,752 samples, and 10% or 33,082 samples for the validation. Two examples in synthetic set are shown in Figure 2.

3. METHODOLOGY

We use deep learning for binary classification to detect volcanic deformation pattern in velocity maps. The expected output is a probability that a patch is containing volcanic deformations. To train such a model, usually a training set with large number of high quality samples is needed. In our case, the number of real velocity maps is limited, and they are reserved for testing purpose, thus the synthetic set which was introduced in Section 2.2 is used to train the CNN. However, this synthetic data often keeps some differences to the real data.

Fine-tuning the model with small hybrid synthetic-real data set (referred as fine-tuning set in this paper) helps to learn the useful patterns in the real data. It is performed by freezing the model except last few layers, and modifying the last few layers using small learning rate with fine-tuning set. This fine-tuning method has been widely used in SAR applications such as image generation (Song et al., 2017). In the following, the process of fine-tuning set generation and comparison of different fine-tuning approaches are presented.

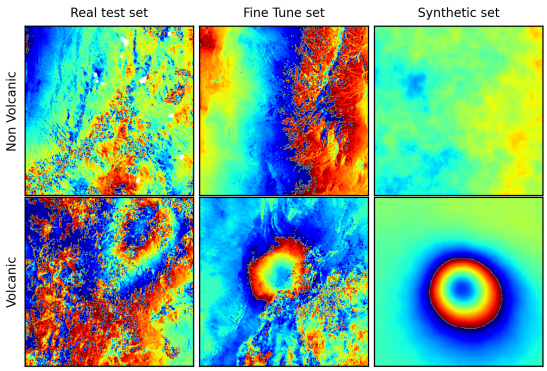

Figure 2. Comparison of the sample images from different data sets. It is visible that synthetic data are missing slope induced signal, and contains only low pass patterns. The fine-tuning set resembles the real data, although at a slightly lower intensity overall. It should be noted that we use simulated volcanic deformations instead of real ones in synthetic and fine-tuning sets.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-3-2022-85-2022 | © Author(s) 2022. CC BY 4.0 License.

3.1 Creation of fine-tuning set

To create it, data were extracted in Andes surrounding the region of interest (mountainous area surrounding tested area shown in Figure 1), but never overlapping it. The closeness to the region of interest insures similarity of the patterns - spatial correlation, but it is noticeable that the patterns extracted around the region of interest are subtler than at the centers of the mountains ridge and volcanoes.

The area extracted this way was split in patches without overlap, checked for missing values, and only data with $<30\%$ missing values has been used. The set was searched through, and examples with large salt lakes in the middle of an image have been removed. This data set contained 836 patches, of slope induced signal with several examples containing salt lakes. There was no volcanic examples in this data set, as all the region with potential volcanic deformations is kept for real test set.

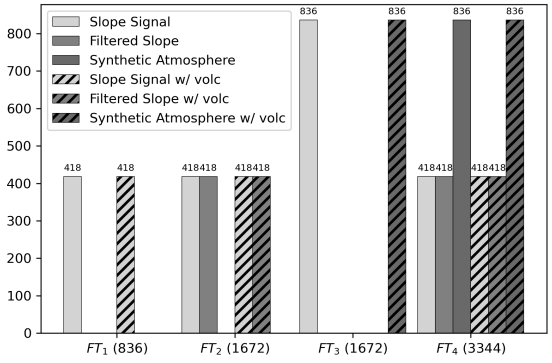

Figure 3. The constituents of the four fine-tuning sets. The slope induced signal set contains patterns extracted from real data in the mountainous region surrounding but not overlapping the region of interest; the filtered slope set uses slope induced signal set with low-pass spatial filtering; and in the synthetic atmospheric set, simulated atmospheric noise is used. All of the volcanic class samples in the sets use simulated volcanic deformations with addition of previous sets.

For fine-tuning, we needed to have the balanced set of two classes: containing volcanic deformations (deformation class) and without volcanic deformations (non deformation class). For that reason, the volcanic simulations had to be included in the fine-tuning set. The visual comparison of constituents of each fine-tuning set can be seen in Figure 3. Four fine-tuning sets have been created and tested:

- FT_1 - Used only slope induced signal data set of 836 samples as the background noise, and half used for deformation and half for non deformation class. Deformation class used 418 randomly selected simulated volcanoes of deformations larger than 1.5 cm added to the slope induced signal and wrapped afterwards and non deformation class used only 418 samples of slope induced signal data set.

- FT_2 - Contained the complete FT_1 set, and added filtered FT_1 using low pass median filter with size of 20 pixels and wrapped afterwards. It consisted of 1672 samples.

- FT_3 - Used complete slope induced signal only as a non deformation class, and for deformation class it uses 836 randomly selected examples of descending track simulated volcanic deformations with simulated residual selected samples of simulated volcanic deformations with noise of descending track (scale of deformations greater than 0.6 cm).This set includes 1672 samples.

- FT_4 - was made using complete FT_2 set and simulated data in equal part. 836 samples of simulated atmospheric noise was randomly extracted as non deformation class, and same amount of simulated volcanic deformations added with simulated atmospheric noise from a descending track. This set totaled 3344 samples.

FT_1 set was created to give a data set without repetition of the background noise. It is the most concise, simplest and straightforward fine-tuning set.

FT_2 had added filtered imagery to the FT_1 set. The filtered set was supposed to be more similar to the synthetic training set, allowing for more generalization of the model, and possibly more smooth training.

In FT_3 we tried using the complete variance from the extracted real background noise set for fine-tuning the non deformation class. This was expected to reduce the number of false positive detections. Since the deformation class includes only simulated residual atmosphere as background noise, this opened a possibility of distribution gap in real data for deformation class.

To increase the number of samples in the fine-tuning set and to make more smooth fine-tuning process FT_4 was created. It includes both synthetic data from training data set and FT_2 set with all the real background noise in equal measure, which was expected to allow for smoothest fine-tuning process and to increase the generalization of the model.

3.2 Approaches to fine-tuning

Four different fine-tuning approaches were tested:

- **[last]** - fine-tuning only last dense layer, while rest of the model is frozen.

- **[-1last]** - Freezing all the layers but second last dense layer, and fine-tuning it for improved high level feature representations.

- **[last2]** - fine-tuning last two layers together while the rest of the model is frozen.

- **[-1&last1]** - Step fine-tuning - first fine-tuning last dense layer while the rest of the model is frozen, then freezing all the layers but second last dense layer, and finally again last dense layer.

Freezing the model and fine-tuning only the last layer is a typical approach which also known as transfer learning (Pan and Yang, 2009). It relies on the model having been thought to identify the significant patterns and features before, which can be used to detect the new classes. It requires the least data among the selected approaches. This approach was used on the model trained on synthetic data to try to teach it to separate between the salt lakes and slope induced signal and volcanic deformations.

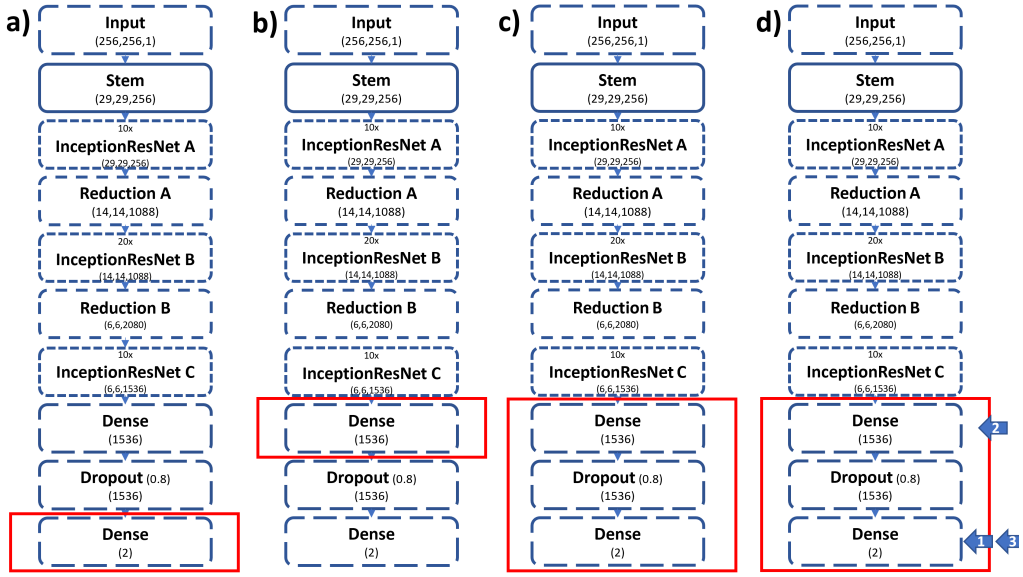

Figure 4. The four fine-tuning approaches on the convolutional neural network architecture. a) [last], b) [-1last], c) [last2], d) [-1&last1]. *Dropout layer does not have any trainable weights, and it works only during training instead of inference stage. Also, it should be noted that every module contains multiple operations

Since we were not completely confident that the synthetic model managed to prepare all the tools to handle the slope induced signal or salt lake patterns from the synthetic data set, approaches which would allow some correction of identified features were tested. First approach to this was only fine-tuning the second last layer so we allow it to improve the high level feature representations, and to feed the final dense layer better input for discrimination. The second approach was fine-tuning the last two layers simultaneously, and third step fine-tuning last two layers independently.

Fine-tuning last two layers should allow some adjustment of extracted high level features as well as correction to how the class probabilities were calculated. Step fine-tuning of last two layers was expected to give slightly better results given the small fine-tuning sets. It would allow the classification layer to adjust to the new data, then in second step to improve understanding of high level features, and finally give the last adjustment to classification using the newly improved features.

4. EXPERIMENT

InceptionResNet v2 architecture (Szegedy et al., 2016) is used as the frontend network. It was selected after comparison of 6 models fully trained on the synthetic set, among which it gave the best performance. The model was trained with early stopping after 10 epochs when validation loss does not decrease any more. The models were always trained to accuracy of 99.9+ % and they also achieved similar results with synthetic test set. On real test set the achieved area under the curve receiver operating characteristic (AUC ROC) value of model was far lower than expected, giving only 57.92 %. After exploring the results using Grad-CAM (Gradient-weighted Class Activation Mapping), it was clear that the model was confused by patterns in real data which have not been accounted in synthetic data (Song et al., 2020): the patterns of salt lakes and slope induced signal. Reducing this gap was performed by fine-tuning of the model on hybrid synthetic-real set.

Two experiments have been performed. First one testing the different approaches on FT_1 fine-tuning set, and second one using the best approach to test the performance of different hybrid sets. The results are compared to the baseline model which was trained on synthetic data without fine-tuning.

As the real test set is highly imbalanced, the accuracy or precision cannot be used as representative metric. While F1 measure is often indicative of overall performance, it is also highly skewed by imbalance. Here the primary goal is to detect all of the volcanic deformations (true positive rate - TPR), and secondary to reduce the number of images that needs to be checked (false positive rate - FPR). The metric which incorporates both of these is receiver operating characteristic (ROC), which is also unaffected by data skew (Jeni et al., 2013). ROC graphs need to be visually analyzed, but area under the curve ROC (AUC ROC) value can be used as numeric representation of the models performance.

The different frame level metrics, depending on how the patch probabilities are merged, can be better suited to some models and therefore misleading. For this reason the patch level metrics are used. Beside the standard metrics, it is also checked if each frame containing a volcanic deformation has at least one patch marking it.

4.1 Effect of different fine-tuning approaches

The four fine-tuning approaches have been tested, and metrics of the models extracted. For each case the base model trained on synthetic data are used for fine-tuning. The model is frozen, only the selected layers have been tuned.

The metrics of the results achieved by different approaches can be seen in the Table 1.

From the results it can be noticed that the TPR rate is reduced in all of the cases in comparison to synthetic baseline model, but with significant improvements in FPR. Since the real test set contains also many partial volcanic deformations, it is a question which should be classified positively and which not. This means that there is a lot of edge cases in the patch level results,

Approach	*TPR*	*FPR*	*AUC ROC*	Accuracy	Precision	F1	Volcanoes
Synthetic	100.0	99.052	55.45	19.097	18.466	31.175	10/10
FT_1_$[last]$	**94.366**	31.754	**86.442**	73.032	40.0	56.184	10/10
FT_1_$[-1last]$	45.07	14.218	75.724	78.323	41.558	43.243	10/10
FT_1_$[last2]$	56.338	**8.057**	84.849	**85.419**	**61.069**	**58.608**	10/10
FT_1_$[-1\&last1]$	0.0	0.0	26.549	81.677	0	0.0	0/10

Table 1. Results achieved on real test set using different fine-tuning approaches.

and that perfect score is hardly reachable. The primary concern of the model is that it detects the whole volcanic deformations, so omission of the few partial deformations or having small number of false positive detections is acceptable.

The first approach, FT_1_$[last]$, to fine-tuning by retraining only the last layer has given the best results according to the TPR and AUC ROC metrics and maintaining the detection of all of the ten frames with presence of volcanic deformations, while reducing FPR by three times, to 31.754 %. If the priority is to keep high TPR on patch level, this would be the best model.

The second approach, FT_1_$[-1last]$, retraining only the second last dense layer has given an improvement in the results in comparison to the baseline but it did not outperform the two best models. This approach was limited by the last fully connected layer, and it was to be expected to be outperformed by other models.

FT_1_$[last2]$ gave many of the best performance metrics, but at a cost of lower TPR. It still maintains the detection of all ten volcanic deformations, while decreasing the FPR almost four times bellow the value of the first approach, or 12 times bellow the baseline model. Here it is easily visible that tuning more layers brings additional benefits. It is reasonable to expect greater improvements with larger fine-tuning sets.

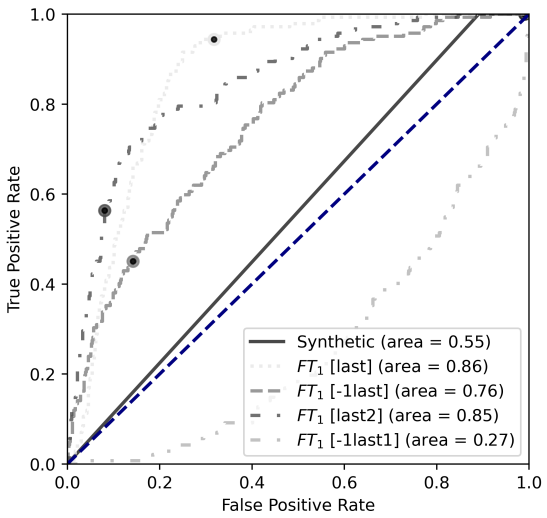

Figure 5. The ROC curves on real test set of the models trained on four tested fine-tuning approaches, and points showing the models thresholds (50% probability of presence of volcanic deformations).

Although both approaches of training individually the last or second last layer, or both brought significant improvements, FT_1_$[-1\&last1]$ approach did not manage to converge to a new optimum by step tuned layers.

The best TPR and AUC ROC result on the smallest fine-tuning set was achieved by the FT_1_$[last]$ approach, which was to be expected. By closer analysis of the ROC curve, in Figure 5, the [last] approach gives the TPR above 90% the at the lowest FPR, although at sub 10 % FPR [last2] model outperforms it.

4.2 Use of different fine-tuning sets

The four different fine-tuning sets have been compared using the best performing approach according to ROC and TPR metrics: [last] - retraining only the last dense layer of the model, while the rest of the model is frozen. This approach demands the smallest size of the data set and therefore is less restrictive to the smaller fine-tuning sets (FT_1).

The metrics extracted from an InceptionResNet v2 model fine-tuned on each of the fine-tuning sets can be seen in Table 2.

Data Set	*TPR*	*FPR*	*AUC ROC*	Accuracy	Precision	F1	Volcanoes
Synthetic	100.0	99.052	55.45	19.097	18.466	31.175	10/10
FT_1_$[last]$	**94.366**	31.754	86.442	73.032	40.0	56.184	10/10
FT_2_$[last]$	74.648	23.065	81.017	76.516	42.063	53.807	9/10
FT_3_$[last]$	0.0	0.0	32.991	81.677	0	0.0	0/10
FT_4_$[last]$	76.056	**14.534**	**88.195**	**83.742**	**54.0**	**63.158**	10/10

Table 2. Results achieved on real test set using the different fine-tuning data sets.

FT_2 set performed worse than FT_1, even though it is double the size and contains complete FT_1 as well as filtered FT_1.

FT_3 has underperformed as we deemed probable. The negative class consisting of slope induced signal was always more similar to the data as it constituted larger parts of the image. Still the balanced, equally split subsets of data are necessary for good fine-tuning performance.

FT_4 gave positive results, because of its more numerous and more diverse samples. FT_2 which is completely used in creation of this set, along with synthetic data, underperformed. The only difference between the FT_2 and FT_4 is the 1672 samples of synthetic data, which on its own did not manage to produce a good model.

By analysing the ROC curve in Figure 6, it is visible that FT_4 performed better by a margin than the FT_1 set in the low FPR region (¡18%), while FT_1 is slightly better suited when higher TPR is desired. FT_4 gave better results than FT_1 in every metric except for TPR, where FT_1 is better by 18%.

With exception of the FT_3 the fine-tuning sets were grouped together, but still show that well prepared fine-tuning set allows for extraction of additional performance. By readjusting the classification threshold, the models can take any of the TPR/FPR performances which their ROC curve presents (for this adjustment a real validation set would be needed), and the best ROC values are given by the highest quality and most concise FT_1 and the largest and most diverse set FT_4.

5. CONCLUSION

In this paper, it is shown how the velocity maps are used with classification DL model to detect subtle volcanic deformations, the challenges created by differences between the synthetic and real data and real data unique patterns and it is demonstrated how the fine-tuning adjusted the DL model to these patterns present in the real velocity maps in order to decrease the model sensitivity to background noise in real data. Due to the limited number of real data, we used simulated velocity maps to initially train the network, and used a hybrid synthetic-real set to fine-tune the last few layers. Different fine-tuning approaches

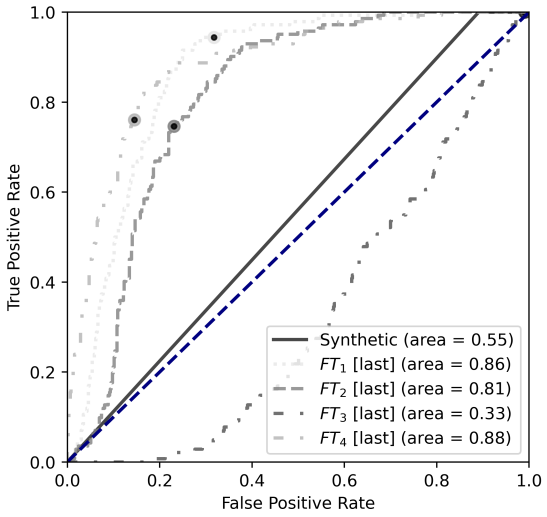

Figure 6. The ROC curves on real test set of the models trained on four tested fine-tuning sets, and points showing the models thresholds (50% probability of presence of volcanic deformations).

and set creation techniques have been tested. Seven fine-tuned models have been created and compared on the important metrics and ROC curve. Results show significant improvements over the initial synthetic approach, up to 12x decrease in FPR and up to 33% increase in AUC ROC. The reduction of FPR came at a cost of TPR, but maintained the detection of all of deforming volcanoes, achieving minimal detected deformation of 0.5 cm/year - an order of magnitude lower than previous work achieved (Anantrasirichai et al., 2019a).

The approach of fine-tuning only the last dense layer and fine-tuning last two dense layers gives the best results. Besides, the fine-tuning set FT_1 which consists only the extracted slope induced signal and deformations, and the set FT_4 which consists of FT_1, filtered FT_1 and synthetic set, allow better performances.

While this work has been performed under assumption that the PSI and InSAR processing of solid Earth tide, tropospheric and ionospheric corrections eliminated the delays in the data completely, it is still present in the regions of very high elevation. It has been shown that none of the tested global weather models accurately captured tropospheric delays in a researched volcano region (Stephens et al., 2020), and that used weather reanalysis (ERA5) performs better in the northern hemisphere, than in southern (Yu et al., 2021). As the elevation is the highest at the ridge of the mountains, along which all of the volcanoes can be found, not many samples containing these patterns have been included in fine-tuning set. This is the main reason why the FPR is not completely minimized by fine-tuning. It is expected that the synthetic and fine-tuning set and model performance can be further improved by accounting for the residual topographic component.

ACKNOWLEDGEMENTS

The paper is based on research performed as a part of TecVolSA project (Montazeri et al., 2021). The velocity map data has been prepared by InSAR team on DLR and special thanks to Robert Shau for patience and help with data. Also great thanks go to GFZ and especially to Rene Mania for simulation of volcanic deformations and Tomas Walter for their help with the domain knowledge on volcano deformations and regional geologic processes. We thank Kun Qian for cordial help with manuscript improvements. For support with their discussions and advice on the DL we thank Yuanyuan Wang, Syed Mohsin Ali, and Ivica Obadić.

REFERENCES

Adam, N., Kampes, B., Eineder, M., Worawattanamateekul, J., Kircher, M., 2003. The development of a scientific permanent scatterer system. *Proceedings of the Joint ISPRS/EARSeL Workshop "High Resolution Mapping from Space 2003"*, ISPRS.

Anantrasirichai, N., Albino, F., Hill, P., Bull, D., Biggs, J., 2018. Detecting volcano deformation in InSAR using deep learning. *arXiv preprint arXiv:1803.00380*.

Anantrasirichai, N., Biggs, J., Albino, F., Bull, D., 2019a. The application of convolutional neural networks to detect slow, sustained deformation in InSAR time series. *Geophysical Research Letters*, 46(21), 11850–11858.

Anantrasirichai, N., Biggs, J., Albino, F., Bull, D., 2019b. A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets. *Remote Sensing of Environment*, 230, 111179.

Ansari, H., De Zan, F., Gomba, G., Bamler, R., 2019. Emi: Efficient temporal phase estimation and its impact on high-precision insar time series analysis. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 270–273.

Bountos, N. I., Papoutsis, I., Michail, D., Anantrasirichai, N., 2021. Self-Supervised Contrastive Learning for Volcanic Unrest Detection. *IEEE Geoscience and Remote Sensing Letters*.

Crosetto, M., Monserrat, O., Cuevas-González, M., Devanthéry, N., Crippa, B., 2016. Persistent scatterer interferometry: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 78–89.

Even, M., Schulz, K., 2018. InSAR deformation analysis with distributed scatterers: A review complemented by new advances. *Remote Sensing*, 10(5), 744.

Ferretti, A., Prati, C., Rocca, F., 2000. Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Transactions on geoscience and remote sensing*, 38(5), 2202–2212.

Gaddes, M., Hooper, A., Bagnardi, M., 2019. Using machine learning to automatically detect volcanic unrest in a time series of interferograms. *Journal of Geophysical Research: Solid Earth*, 124(11), 12304–12322.

Gaddes, M., Hooper, A., Bagnardi, M., Inman, H., Albino, F., 2018. Blind signal separation methods for InSAR: The potential to automatically detect and monitor signals of volcanic deformation. *Journal of Geophysical Research: Solid Earth*, 123(11), 10–226.

Gaddes, M., Hoopera, A., Albinob, F., 2021. Simultaneous classification and location of volcanic deformation in SAR interferograms using deep learning and the VolcNet database.

Jeni, L. A., Cohn, J. F., De La Torre, F., 2013. Facing imbalanced data–recommendations for the use of performance metrics. *2013 Humaine association conference on affective computing and intelligent interaction*, IEEE, 245–251.

Montazeri, S., Ansari, H., De Zan, F., Mania, R., Shau, R., Beker, T., Parizzi, A., Haghshenas Haghighi, M., Niemz, P., Cesca, S. et al., 2021. Tecvolsa: Insar and machine learning for surface displacement monitoring in south america. *EGU General Assembly Conference Abstracts*, EGU21–6086.

Nikkhoo, M., Walter, T. R., Lundgren, P. R., Prats-Iraola, P., 2016. Compound dislocation models (CDMs) for volcano deformation analyses. *Geophysical Journal International*, ggw427.

Pan, S. J., Yang, Q., 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345–1359.

Parizzi, A., De Zan, F., Gonzalez, F. R., Ansari, H., Gomba, G., Brcic, R., Eineder, M., 2019. Processing and performance of large scale deformation monitoring with insar stacks. *Geophysical Research Abstracts*, 21.

Song, Q., Chen, H., Xu, F., Cui, T. J., 2020. EM Simulation-Aided Zero-Shot Learning for SAR Automatic Target Recognition. *IEEE Geoscience and Remote Sensing Letters*, 17(6), 1092-1096.

Song, Q., Xu, F., Jin, Y.-Q., 2017. Radar image colorization: Converting single-polarization to fully polarimetric using deep neural networks. *IEEE Access*, 6, 1647–1661.

Stephens, K. J., Wauthier, C., Bussard, R. C., Higgins, M., LaFemina, P. C., 2020. Assessment of mitigation strategies for tropospheric phase contributions to InSAR time-series datasets over two Nicaraguan volcanoes. *Remote Sensing*, 12(5), 782.

Sun, J., Wauthier, C., Stephens, K., Gervais, M., Cervone, G., La Femina, P., Higgins, M., 2020. Automatic detection of volcanic surface deformation using deep learning. *Journal of Geophysical Research: Solid Earth*, 125(9), e2020JB019840.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.

Valade, S., Ley, A., Massimetti, F., D'Hondt, O., Laiolo, M., Coppola, D., Loibl, D., Hellwich, O., Walter, T. R., 2019. Towards global volcano monitoring using multisensor sentinel missions and artificial intelligence: The MOUNTS monitoring system. *Remote Sensing*, 11(13), 1528.

Yu, C., Li, Z., Blewitt, G., 2021. Global comparisons of ERA5 and the operational HRES tropospheric delay and water vapor products with GPS and MODIS. *Earth and Space Science*, 8(5), e2020EA001417.