

Abnormal Change Detection in Volcano SAR-Images using Artificial Neural Networks

Luibrand David¹, Erik Gruener¹

¹*Computer Vision and Remote Sensing, Institut for technical computer science and micro electronic,
Technical University of Berlin, Straße des 17. Juni 135, Berlin, Germany
{luibrand, erik.gruener}@campus.tu-berlin.de}*

Keywords: Abnormal Change Detection, Image Prediction, Autoencoder, SAR, Volcano

Abstract: Monitoring a huge amount of geological changes in a natural environment is a task prone to be solved by computer vision. A very reliable source of images for such a task come from radar images periodically taken by satellites. However natural changes such as rainfall or vegetation growth have significant impact on these images. In this paper we train an artificial neural network to detect such changes and predict how a radar image should look given these changes. By computing a distance to the real image, we can tell, if an abnormal change happened. This is done on images of volcanoes and without the need of labeled data.

1 INTRODUCTION

Monitoring volcanoes for warning signs, specifically significant structural change, that might tell of upcoming or ongoing volcanic activity, can save lives. With modern technological advancements such as satellites equipped with radar, it has become possible to monitor the approximately 1500 currently active volcanoes on earth [Smithsonian Institution, 2013], no matter how remote they are nor the weather condition on earth. Radar is a very powerful tool in that it is not influenced by illumination nor clouds, which means radar images can be taken at any time of day. The downside on the other hand is the speckle noise in the images and changes in reflection at different time stamps, due to different surface reflection-properties, such as wet or vegetation areas. This change is reflection, between two time instances, leads to inaccurate outcomes with the usual change-detection algorithms, which is why we propose a new, more robust approach which neglects normal change. Additional factors that enhance the models performance are the additional images from different orbits and angles of the volcanoes. An example of an image pair from the data set, used in the models, is shown in Figure 1.

Analyzing satellite images has the potential to distinguish abnormal changes, that could be a warning sign for a coming eruption or change in an ongoing eruption, from normal seasonal or environmental changes that are not caused by volcanic activity. We propose a novel approach to increase change de-

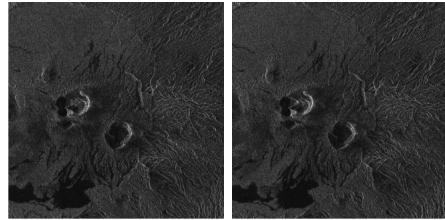


Figure 1: Image pair example of the volcano: Ambrym. The left image is the SAR VV intensity image at date: 04/04/2019 and on the right side date: 16/04/2019

tention accuracy and automation by leveraging big data and deep learning methods. A data set images of 17 volcanoes are used to train and test a deep convolutional-autoencoder (CAE). The autoencoder technique is quite promising for creating future normal predictions of past images that can then be compared to possibly abnormally changed present images. This approach does not require any labeled data, and should be able to overcome noise and other differences between the images, in order to provide an efficient method for abnormal change detection in volcano images.

Two slightly different approaches are tested by separate teams. The first approach uses 128x128 pixel patches from the past in addition to 3 randomly chosen 64x64 patches from the present image, also called 'Global context' in this manuscript, as input into the Network. The output of the ResNet [He et al., 2015] like CAE model, also referenced as the prediction, is

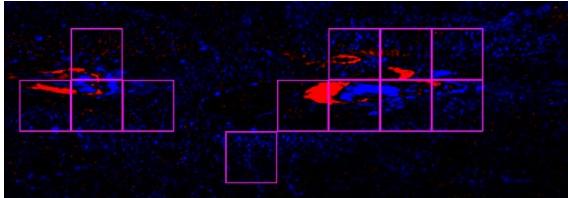


Figure 2: General idea of the model architecture. Day 1 being the past volcano image and Day 6 representing the present image. The blue vertical lines indicating the encoding and decoding process and the green squares representing the added global context information.

then compared to the patch from the present image. The second model follows a more holistic approach of feeding the entire past image together with Global context patches into the CAE. The prediction is then again compared to the current image to receive an error of difference between the images. The general architecture of this approach of informed prediction can be seen in Figure 2.

We are confident to say that this novel approach is a promising start into using autoencoder and other deep NN’s in the change detection field for SAR images.

2 RELATED WORK

Since the upcoming of convolutional neural networks (CNN) in 2012 many achievements have been made in the field of computer vision. The first artificial neural network (ANN) to significantly outperform other state of the art algorithms in image recognition was AlexNet [Krizhevsky et al., 2012], which won the ILSVRC in 2012 [Alom et al., 2018]. The most recent architecture of CNNs to performed best at the ILSVRC 2015 are the once using residual layers [He et al., 2015].

Another version of neural networks are autoencoders. They can be used, for instance, to make predictions on movements [Walker et al., 2016] or to denoise images [Anantrasirichai et al., 2018a].

Change detection also got a boost in accuracy due to ANNs [Li et al., 2019, Huang et al., 2017, Anantrasirichai et al., 2018b].

As labeled data are usually scarce, CNNs often get pretrained, for instance on a large data set of natural images [Anantrasirichai et al., 2018b] or by training an autoencoder with SAR images and then detaching the decoder [Huang et al., 2017].

In order to increase the amount of collected data, augmenting them usually by rotation [Anantrasirichai et al., 2018b] or by translation and mirroring [Huang

et al., 2017] is often used.

To obtain labeled data Li et al. [Li et al., 2019] and Gong et al. [Gong et al., 2016] use a spatial fuzzy clustering and a fuzzy c-mean clustering respectively in order to get labeled pixels. This data is then used to train their CNN. They use a two-channel input for their CNN one channel containing the picture before and the other after a potential change. Both Li et al. [Li et al., 2019] and Gong et al. [Gong et al., 2016] use a very small input layer of 7×7 and 5×5 to classify small areas.

On top of Gong et al. [Gong et al., 2016] Dong et al. [Dong et al., 2019] use a pseudo Siamese network where the convolution of both channels is independent, this information is then fused in the later fully connected layers in order to determine if there are changes.

Anantrasirichai et al. [Anantrasirichai et al., 2018b] learns one specific change (earth movement), thus their CNN only needs one picture as input. They obtain labeled data by labeling changed areas of a recorded eruption of a volcano. To additionally reduce noise a Gaussian filter is applied to every image. Their trained CNN then assigns a probability of change to overlapping areas in the image. To develop an alert system, they use the highest probability of a change of all patches and a threshold.

No work was found that does not need labeled data nor one that makes predictions given certain changes.

3 DATA SET

The Sentinel-1 mission provides high quality C-band SAR data, acquired by twin satellites, Sentinel-1A and Sentinel-1B in an 6 days interval. Our data-set contains 1512 SAR images from 17 volcanoes(135GB), containing VV and VH polarized data. The images are not always aligned and come in different dimensions ranging from $(800,3000)$ to $(1000,4000)$. The image dimensions are the same though for the same volcano in the same orbit. The raw .tif files include mostly 4 channels which are then further processed using the log-intensity of the VV and VH polarizations for better representation and computation efficiency.

Having 1512 images with 4 channels each from 17 volcanoes 3 totaling 42 orbits, the amount of pairs that fit together as of (past,present) is decreased to 1471 images. Unlike other change detection methods [Radke et al., 2005] we do not first construct a difference image and apply some model as is usual in postcomparison methods [Gong et al., 2015], but to simplify the process by using the image data without



Figure 3: Global locations of volcanoes in the used data set.

introducing algorithms that might distort the information.

Two different architectures are being tested and were developed independent of each other, which is why further data processing will be presented in the separate sections. This includes the patch creation of the first model as well as the different approaches to generating the training and testing data-sets.

4 METHOD 1

This section describes our first approach.

4.1 Data Preparation

In this sub section we first describe how we normalized our data and then how we created our training set.

4.1.1 Normalization

Each image we obtain comes with two complex values representing the horizontal and vertical polarization. We create the magnitude for both polarization and obtain the vertical intensity image VV and horizontal intensity image VH. This way we reduce the input dimension and still preserve important information of both images. In addition, structures in intensity images are better recognizable when looked at by humans.

As intensity images still contain values between $[0, \sim 50k]$ we take the logarithm for each pixel plus 1, in order scale the values in a less intensive way. We add 1 in order to obtain only positive numbers.

As shown by Ioffe and Szegedy [Ioffe and Szegedy, 2015] neural networks can learn more efficient when their input belongs to one distribution, thus we normalize to a standard normal distribution using the mean and standard deviation std.dev of 100 randomly picked images for VV and VH respectively. We then stack VV and VH images to form

two-channel images. In short our normalized input will be

$$(\text{norm}(\text{VV}, \cdot, \cdot), \text{norm}(\text{VH}, \cdot, \cdot))$$

with

$$\text{norm}(\text{pol}, x, y) = \frac{\log(\text{pol}_{x,y} + 1) - \text{mean}_{\text{pol}}}{\text{std_dev}_{\text{pol}}}$$

For simplicity we will now omit the third dimension and treat it like a single channeled image.

4.1.2 Creating the Data Set

As described pictures come in pairs showing the same cutout of one volcano. We refer to the pictures as *interest* and *future* for the picture of day t and day $t + 6$ respectively. We crop corresponding areas of size 128x128 in both images and obtain crop pairs (*interest_crop*, *future_crop*) (red area in Figure 4). We do so in a grid and random based manner. We assign each pair a set of three global context picture pairs. Each of them consisting of two crops of size 64x64 randomly taken from the border of *interest* and *future*. We refer to them as *gc_pre_i* and *gc_post_i*, where *gc_pre_i* is taken from *interest* and *gc_post_i* is taken from *future* (green areas in Figure 4). The index i denotes the index of one of the three crops we take. In total we obtain around 135.000 training samples, which we shuffle randomly.

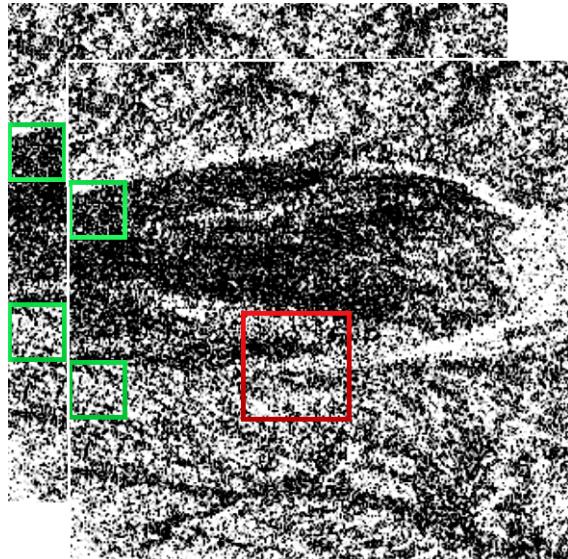


Figure 4: Example of one data set. Red denotes the interest area (and the future area behind it), green denotes the global context.

We split our data set into one for training containing 452 image pairs (92%) and one for validation containing 36 image pairs (8%). The validation set con-

tains time intervals for which we know, that the volcano was active.

4.2 Architecture

Our neural network consists of four main parts *encoder*, *global context*, *fusion* and *decoder*. We describe them and their purpose in this section. Figure 5 shows a simplified version of the network. The entire architecture can be found in Figure 10.

The *global context* has the purpose of providing information about the general change (like growth of vegetation or rainfall) that happened between day t and day $t + 6$ (referred to as day 0 and day 6 in Figure 10). To compute this, a standard CNN with five convolution plus max pooling layers and three fully connected layers is applied to each gc_pre_i and gc_post_i individually using shared weights. This results in a feature vector of size 64 for each input. We then subtract the feature vector of gc_post_i from gc_pre_i which results in a vector telling the difference from *interest* to *future*. We repeat this 3 times during training and 16 times for evaluation. We then take the maximum of each feature. This helps to get information from all different kinds of terrain.

The *encoder* reduces the dimension of the *interest* image in order to force it to ignore noise and speckles and focus on the underlying structure of the image such as water or wood. It consists of three blocks of residuals from ResNet [He et al., 2015] following a max pooling. This results in a 16x16x18 matrix. This was empirically chosen by changing the depth of the architecture (and thus the size of our bottleneck), training each version with 5000 iterations and compare the output with each other. We increased the size of the bottleneck until the smoothness did not change significantly for a human anymore. The size of the chosen bottleneck has 6.25% of the original input. Every pixel sees a field of approximately 35x35 of the original input.

The *fusion* is the part, fusing the *global context* with the output of the *encoder*. To do so we stack the result of the *global context* on top of each pixel of the bottleneck, resulting in a matrix of size 16x16x(18 + 64). Then we apply two (1,1,82) convolutions to fuse the information of the results of the *encoder* and the *global context*.

Then follows the *decoder*, which is built exactly like the *encoder* but with upsampling instead of max pooling. This leads to the final output which is referred to as *prediction* in Figure 5.

During training the loss is computed as the mean squared error (mse) between the output of the *decoder* and the *future_crop*.

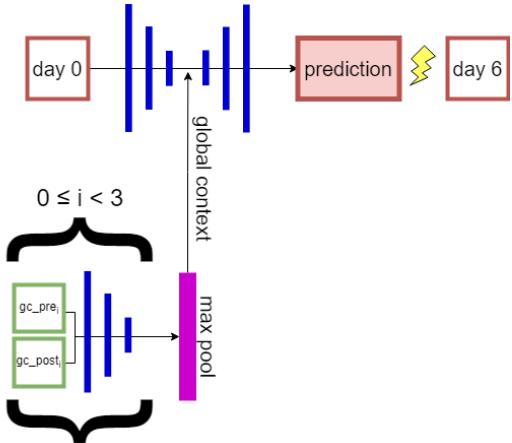


Figure 5: Simplified version of the architecture. The lightning indicates the true value that the prediction should have.

4.3 Evaluation

For evaluating our trained network we reshape the architecture in order to increase the amount of the used *global context* from 3 to 16. This makes the *global context* more reliable as it lowers the likelihood of seeing only certain landscapes (i.e. only grass or only rocks).

We then apply our network in the described manner to a given area of an unknown image. To make it comparable to a *future* image, we also apply our network to the *future* image, but only use one global context with input gc_post_0 . This will cause to *global context* to be $\vec{0}$ and thus the *decoder* will not apply a context change to the *future* image. This has the benefits of having a smoothed version of both images and that both images follow the same distribution. We then compute the mse for the predicted area and it's *predicted future* area. If the mse crosses an empirically chosen threshold, an abnormal change was detected. The pipeline of this can be seen in Figure 6.

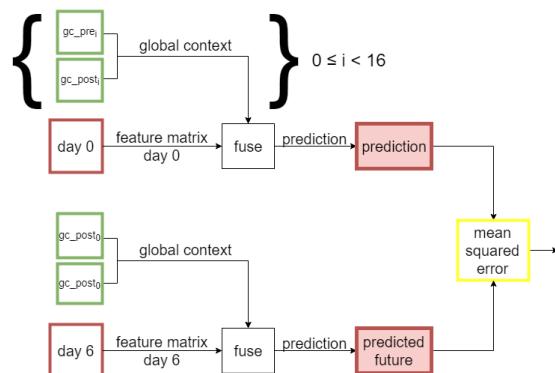


Figure 6: Pipeline of the evaluation.

5 DISCUSSION - METHOD 1

In this section we will discuss results of our first approach. As ground truth, given by visible light images, is only available sometimes, we have to compare our images qualitatively. We obtain these information of Sentinel-Hub [Sinergise, 2017]. A direct comparison to other work is not possible as our topic of detection abnormal changes in a big set of SAR images is new in the field of computer vision.

The mse for the training data is approximately 0.58 and 0.6 for the validation data. As the difference is insignificant and likely caused by eruptions, we conclude that overfitting was avoided.

A comparison of a predicted images and the real SAR images can be seen in Table 3. Our approach detected every area and only the areas where the crater changed. When running the same experiment with raw SAR images, loads of random areas get detected. Yet the image c) is not as bright as image d). We looked at other examples where brightness changed and found, that our *global context* (gc) only worked in few cases. The impact of the discussed in Section 5.1. We conclude, that our algorithm smooths images, such that noise and speckles are eliminated, which still helps to find abnormal changes.

Another abnormal change that was detected can be seen in Table 1. The visible image, taken between the recording of the interest and future image, shows an eruption.

A common source of false positives is the presence of water, as waves and wind cause random differences in water density. Examples of this can be seen in Table 2. As we used these images for training as well, we assume that our trained model would be better, if we would have ignored patches with water. Classifying water should be an easy task, which we leave to future work.

5.1 Impact of the Global Context

Like in Table 3, other images showed no impact of an environmental change. Thus, we ran multiple tests to find out, if the information, provided by the gc, improved the accuracy of our prediction and found that it had an insignificant impact on the loss.

The average absolute weights for every layer that belongs to the gc are above 0.03, most of them are above 0.1. A complete disregard of this part can thus not be assumed.

We qualitatively compared the results of images produced by our network with and without the use of the gc and found that it seems to work for a few predictions. The example of Table 4 shows two such

Table 1: Abnormal change detection for volcano Krakatau on 05/09/2018 and its visible light counterpart.

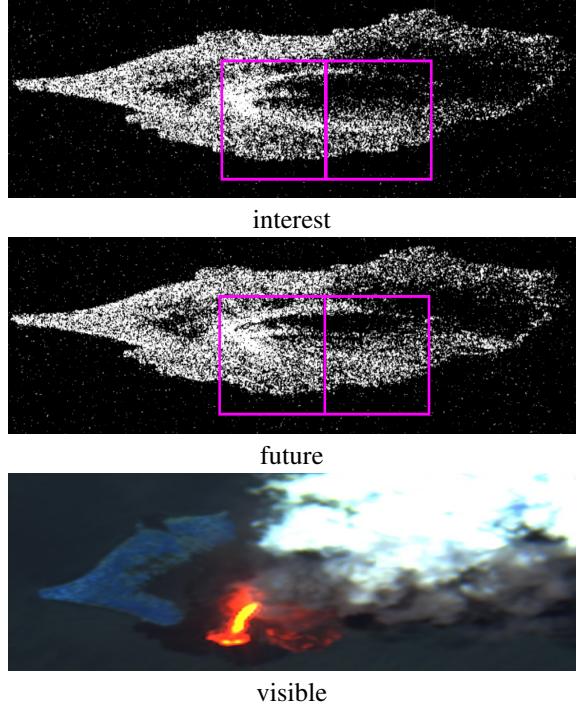
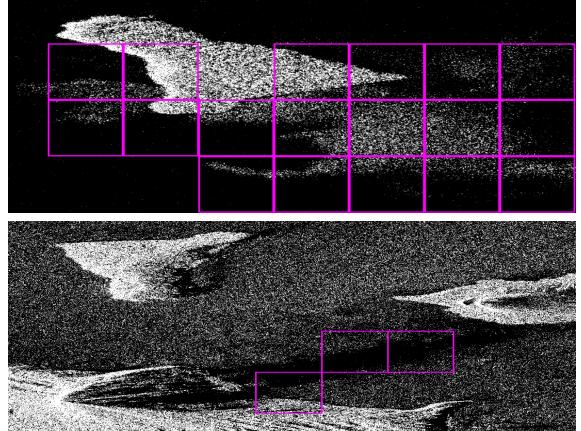


Table 2: Abnormal change detected in water.



images. The future image is a lot brighter, yet only the top middle part of the image with gc shows an increased brightness.

As it effects the entire crop and the perceptive field of one pixel in the bottleneck is limited, this change could only have come from the gc, as we would see more fine-grained changes otherwise. For future work we suggest experiments, for which only one big gc picture shrunk to a normal size should be used. This eliminates the randomness that comes with taking random crops for the gc. If this picture is

Table 3: Comparison of SAR and predicted images of a volcano that erupted.

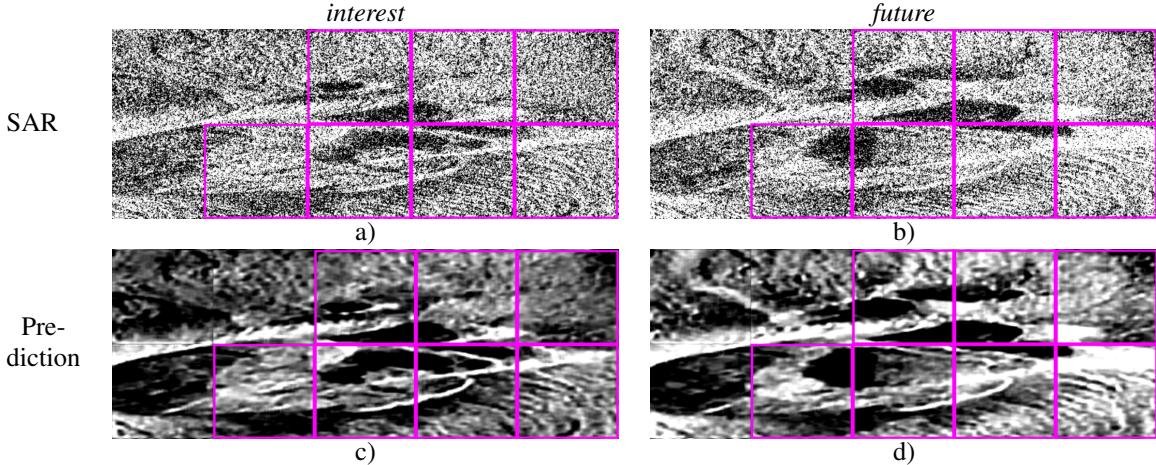
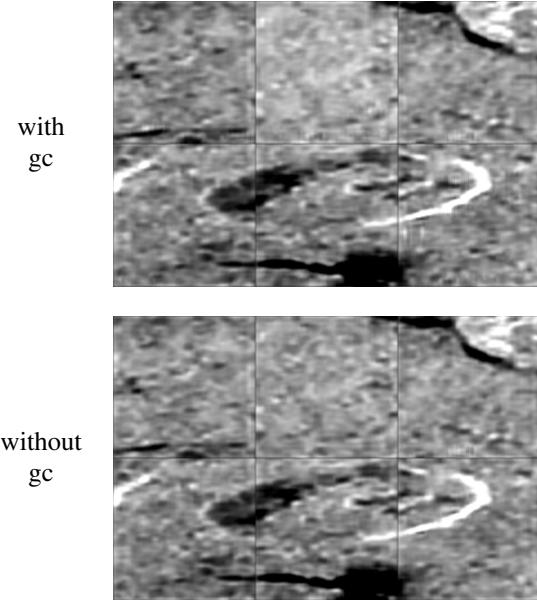


Table 4: The same cutout of a predicted image. The gc is set to 0 for the bottom one. The top middle prediction of the image with gc is brighter than the rest.



chosen the right way, a full view for every possible terrain could be achieved.

We also tried to train our ANN with and without the use of a gc and found, that both versions converged to the same loss value after the same amount of iterations (50k iterations, converged to around 0.64).

As mentioned, the weights of the gc part are not 0, yet it is possible, that, while in training, these weights get set to an insignificant value as it could be more beneficial to find proper weights for the autoencoder part first. At the point where all weights for the

autoencoder are close to convergence, the gc weights could be too small to leave their local minimum with the now small learning rate. This could be avoided by using an L2 weight regularization such that it is beneficial to keep the gc part significant. Alternatively, the training could be restarted with previously learned weights but reseeded weights for the gc part. This way, it could be more beneficial to change weights for the gc part but leave the other values the way they are.

Another possibility is that our encoder is not deep enough to detect underlying structures and thus focuses on compressing the data. To increase the potential of detecting structures, the depth of the encoder needs to be increased. To avoid focusing on compression, the size of the bottleneck needs to be increased. Though this could lead to overfitting, we can easily increase the amount of training data in order to avoid this.

6 METHOD 2

This section describes the second approach, which takes on a more holistic approach to this problem by inputting not just patches and working on a small scale but rather importing the entire image with all its complexity with an additional information from the present image. This has the downside that the data set will be much smaller and training time increases due to the amount of data to process.

6.1 Data Preparation

As stated previously, the networks are trained on the intensity images calculated from the channels of the raw .tif data, which decreases the data set already in half. The training set consist of 1.550 randomly shuffled pairs, with the exception of some volcanoes due to inconsistency and distorted data, and the test set is chosen to be the data from the volcanoes Ambrym, Fuego and Bezymianny. Due to issues with the large amount of data on a personal computer, the size of the past images was decreased to 512x512 pixels. This does not mean that they were cropped, but merely resized to make them easier for processing and unify the sizes, as they may vary for different volcanoes and the CNN does have a static input size it uses. The previously mentioned global context has been chosen to be a patch of size 128x128 taken from the present image. The images are not smoothed or modified in any further way for now, to test the raw outcome. Further improvement can then be tested after the basis has been established.

6.2 Architecture

Inspired by the architecture of the AlexNet [Krizhevsky et al., 2012], this model incorporates the convolutional encoding part and adds a decoder which consist of the deconvolution from the bottle neck. With this setup a deep autoencoder is established and theoretically functional to perform predictions of the input image.

Having humans in mind, and the way humans can improve their predictions, additional information is added to the network. If a human is given a stating image and requested to predict the image six days in the future, it can be easily done, but the accuracy will not be too great if environmental changes are not thought of. Telling the person that it will only rain next week, the prediction will include this additional information to make a more educated guess. This process is included in the NN model by including a patch from the future scenery.

The architecture consist of 4 convolutional layers and 4 deconvolutional layers as well as an adding layer in between to add the two convoluted inputs. As in the AlexNet the convolutions have filter sizes 96,256,384,384 respectively and the same in reverse for the deconvolution, additionally the filter sizes of 11x11, 5x5, 3x3, 3x3 are used with the padding set to same. Starting out with a big filter size the model focuses on larger area changes and with the smaller sizes it will search in smaller areas for patterns. Using the MaxPooling function, the input of size 512x512 is

compressed to a 64x64 image, which forces the model to only learn important features of the original to store it in a compact way. With the padding set to 'same' the matrix stays the same size after the convolution.

As an loss function the mean-squared-error has been chosen, as it was used other image recognition and previous change detection algorithms as well. The activation function for the neurons is the rectified linear unit (relu) as suggested by [Chigozie Enyinna Nwankpa and Marshall, 2018].

6.3 Evaluation

Because all of the data is unlabeled in terms of significant volcanic activity there is no separate data set to test the model on. As a solution three volcanoes are chosen to be tested on. For these the volcanic activity was analyzed to know the dates of volcanic activity. These dates can then be cross-referenced with the model output. Given the created evaluation set there are two steps that are taken. First the image pair is fed into the network to receive a prediction. Secondly this prediction is compared to the original present image that is also run through the CAE to remove speckle and make it comparable. This comparison is currently done with the loss function mean-squared-error. A threshold is manually set after the histogram analysis and is then used in the evaluation process to flag only pairs that are above this value.

7 RESULTS AND DISCUSSION - METHOD 2

In the following section the results of the holistic abnormal change detection approach are presented and discussed.

To see if abnormal changes have occurred in the images of the Ambrym volcano, the evaluation set was applied to the created model and the output prediction was then checked on the current image. The results are visible in 8 and show that for the 70 images of both VV, VH polarization in orbits 59 and 91 only 2 image pairs seem to have abnormal changes detected. This result does not match to the previously, manually located, image pairs that have abnormal change, therefore this model is not ready for practical use.

As a baseline to compare the results from the trained model the mean squared error between the past and present image are computed. The baseline graph for each of the three selected volcanoes can be viewed in Figure7. With the volcanoes data on volcanic activity compared to the data of the baseline it is shown that the baseline graph has a high false positive

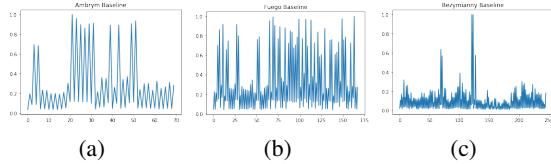


Figure 7: (a) Ambrym Baseline (b) Fuego Baseline (c) Bezymianny Baseline. The y-axis represents the Error and the x-axis the sample pair.

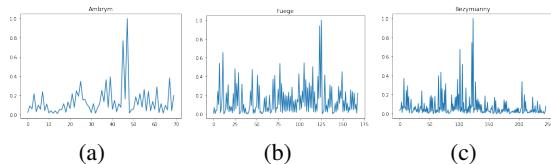


Figure 8: Results of the different evaluation sets used in the trained model. (a) Ambrym CAE (b) Fuego CAE (c) Bezymianny CAE. The y-axis represents the Error and the x-axis the sample pair.

rate. For the Fuego volcano almost all days have similar change in them and volcano Bezymianny show only 2 major peaks that indicate major change. The evaluation set is fed into the trained CAE. The result of this experiment can be seen in Figure 8. Compared to the baseline the model learns to distinguish some characteristics more than others, but what it is precisely learning is not known yet.

Areas that potentially limit the results of the model are the data preparation process, the model design and the evaluation technique. The holistic followed approach did modify the original data quite a bit by forcing the images into a squared shape well below their dimensions. This does have the potential to distort the information in it. When using patches of images they do not have to be resized, but do lack the entirety of the contained information. Due to the same augmentation of the two image pairs, which do come from the same volcano and orbit and therefore have the same dimensions, the distortion is equivalent in both images.

The design of the model is rather simplistic and was designed with the intention of serving as a proof of concept for the idea of global context and deep learning applied to SAR change detection of volcanoes. More sophisticated ways such as the use of a residual network, drop out methods and the binary classification technique in already discussed papers, can be applied to make the model faster and better so that it can truly be used in a practical way.

Working with unlabelled data makes it somewhat difficult to evaluate the model, because no ground truth is available to test the model against. The technique used in model 2 is an imprecise and manual way of labelling and creating a testing set, but can

be a reliable source if done by experts. This process would require not much time from the experts to label the date of volcanic activity for a volcano. Other papers working on SAR change detection work with disaster data sets. Although the specifics are little bit different, this model could also be applied to those data sets, as they do have ground truth data and are partially labeled. The evaluation in model 2 shows that on dates with volcanic activity has occurred the change is detected and on other dates where environmental changes have happened the model learned to neglect the normal changes to some extend.

8 CONCLUSIONS

The process of change detection in volcanic SAR imagery can be optimized and made less labor intensive with the use of machine learning methods such as a convolutional-autoencoder. The presented approaches give promising reasons for further research in the use of deep learning models with SAR images and the technique of adding additional information to the models decoder. Where other methods take the difference at the beginning of their algorithm, our models work with the raw data and apply the difference at the end between the 'educated guess' prediction and the actual present image. While the output images look quite promising in model 1, modification can be done in the data preparation as well as the evaluation process with more sophisticated methods than the mean-squared error and more specific image selection and preparation.

Because the data set in unlabeled there has not yet been any decisive results of how many actual changes have been detected or have not been picked up, but this will be done in the next step to provide some concrete result that other algorithms can be compared on.

The models one and two were tested with 70,000 and 2,400 images respectively. The goal was to create a model that can recognize unnatural difference no matter which orbit and view of the scenery. This goal was not fully met, but a good first step has been done for further research.

This idea of adding additional knowledge into change detection might not just improve methods in the volcanic field, but can be used in the various other change detection areas mentioned in [Radke et al., 2005]

ACKNOWLEDGEMENTS

The main idea of this paper comes from Ley, Andreas, who also gave valuable feedback to us.
We also thank mounts_project [Valade, 2017] for the data and geological insights into volcanoes.
We thank Davis, James for providing clean data.

REFERENCES

- Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Hasan, M., Esesn, B. C. V., Awwal, A. A. S., and Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches. *CoRR*, abs/1803.01164.
- Anantrasirichai, N., Biggs, J., Albino, F., Hill, P., and Bull, D. (2018a). Application of machine learning to classification of volcanic deformation in routinely generated insar data. *Journal of Geophysical Research Solid Earth*, 123(8):6592–6606.
- Anantrasirichai, N., Biggs, J., Albino, F., Hill, P., and Bull, D. (2018b). Application of machine learning to classification of volcanic deformation in routinely generated insar data. *Journal of Geophysical Research: Solid Earth*, 123(8):6592–6606.
- Chigozie Enyinna Nwankpa, Winifred Ijomah, A. G. and Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning.
- Dong, H., Ma, W., Wu, Y., Gong, M., and Jiao, L. (2019). Local descriptor learning for change detection in synthetic aperture radar images via convolutional neural networks. *IEEE Access*, 7:15389–15403.
- Gong, M., Zhao, J., Liu, J., Miao, Q., and Jiao, L. (2015). Change detection in synthetic aperture radar images based on deep neural networks. *IEEE transactions on neural networks and learning systems*, 27.
- Gong, M., Zhao, J., Liu, J., Miao, Q., and Jiao, L. (2016). Change detection in synthetic aperture radar images based on deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 27(1):125–138.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385.
- Huang, Z., Pan, Z., and Lei, B. (2017). Transfer learning with deep convolutional neural network for sar target classification with limited labeled data. *Remote Sensing*, 9:907.
- Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.
- Li, Y., Peng, C., Chen, Y., Jiao, L., Zhou, L., and Shang, R. (2019). A deep learning method for change detection in synthetic aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–13.
- Radke, R., Andra, S., Al-Kofahi, O., and Roysam, B. (2005). Image change detection algorithms: A systematic survey. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 14:294–307.

- Sinergise (2017). sentinel-hub. <https://www.sentinel-hub.com/>.
- Smithsonian Institution, Venzke, E. (2013). Volcanoes of the world. http://www.esa.int/Our_Activities/Observing_the_Earth/Space_for_our_climate/Volcanoes. <https://doi.org/10.5479/si.GVP.VOTW4-2013> v. 4.8.0., Accessed: 20.07.2019.
- Valade, S. (2017). mounts-project. <http://mounts-project.com>.
- Walker, J., Doersch, C., Gupta, A., and Hebert, M. (2016). An uncertain future: Forecasting from static images using variational autoencoders. *CoRR*, abs/1606.07873.

APPENDIX

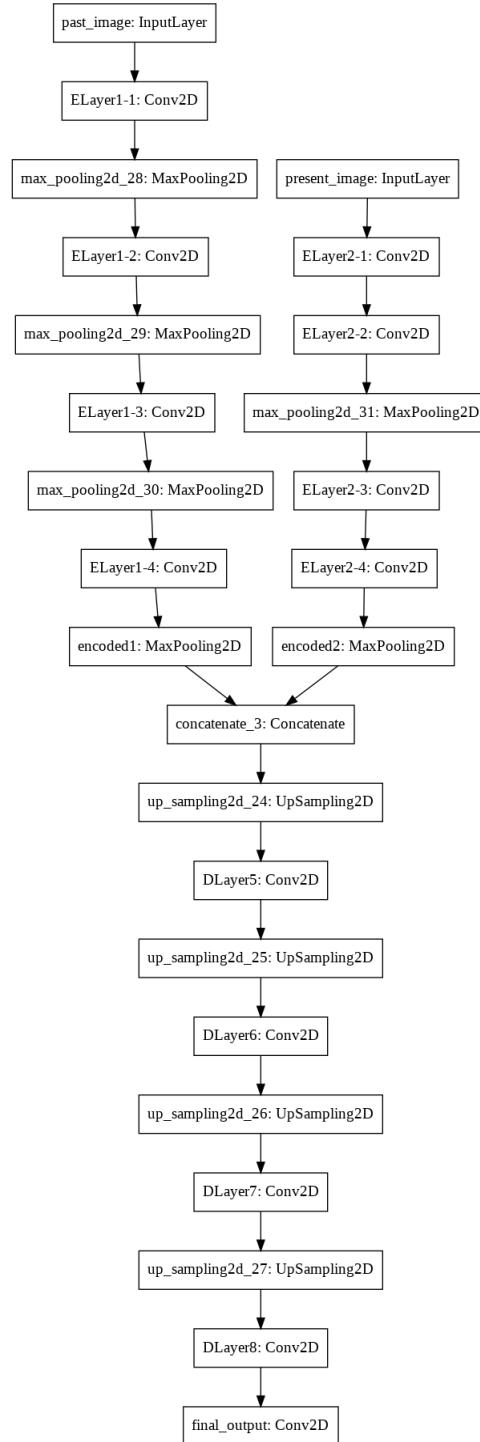


Figure 9: This figure shows the complete architecture of the model. As described in 6 the input consist of a 512x512 pixel image of the past image and an 128x128 pixel crop of the present image. The output is again a 512x512 image that is then used for evaluation with the actual current state.

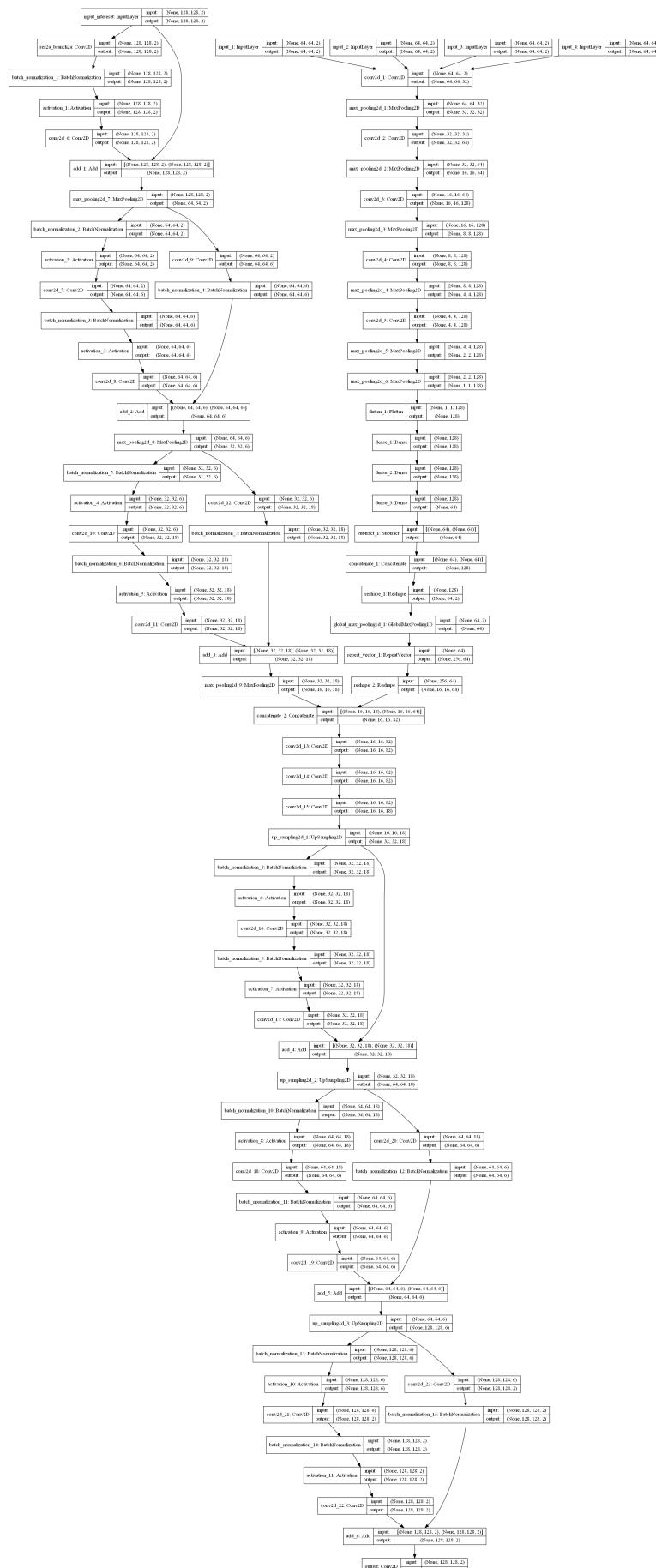


Figure 10: Detailed model of architecture of method 1. Note that this is the training and not the evaluation version. Also note that the shared weights are not represented. Left top is the autoencoder, right top the global context with two global context inputs.