ARTIFICIAL GENERATION OF BIG DATA FOR IMPROVING IMAGE CLASSIFICATION: A GENERATIVE ADVERSARIAL NETWORK APPROACH ON SAR DATA

*Dimitrios Marmanis[1,3], *Wei Yao[1], Fathalrahman Adam[1], Mihai Datcu[1], Peter Reinartz[1], Konrad Schindler[2], Jan Dirk Wegner[2], Uwe Stilla[3]

[1]Department of Photogrammetry & Image Analysis, German Aerospace Center (DLR), Germany
[2]Photogrammetry & Remote Sensing Group, ETH Zurich, Switzerland
[3]Department Photogrammetry & Remote Sensing, Technische Universitaet Muenchen (TUM), Germany

ABSTRACT

Very High Spatial Resolution (VHSR) large-scale *SAR* image databases are still an unresolved issue in the Remote Sensing field. In this work, we propose such a dataset and use it to explore patch-based classification in urban and peri-urban areas, considering 7 distinct semantic classes. In this context, we investigate the accuracy of large CNN classification models and pre-trained networks for SAR imaging systems. Furthermore, we propose a Generative Adversarial Network (GAN) for SAR image generation and test, whether the synthetic data can actually improve classification accuracy.

***Index Terms*—** Big Data, SAR classification, GANs, Generative Adversarial Networks, Deep Learning

1. INTRODUCTION

Classification of very high resolution (VHR) *SAR* image data remains a hard and time-consuming task. Major difficulties include the scarcity of available data, and the challenge of semantically interpreting the *SAR* backscatter signal. Linked to those difficulties, there are no large-scale, *SAR-derived* image databases for Remote Sensing image analysis and knowledge discovery. Furthermore, while optical image classification has seen a breakthrough with the advent of *Deep Learning* methods that require Big Data, SAR-based systems have so far not experienced the same progress, likely because of not enough data with associated training labels is available.

In this work we try to tackle the lack of training data, by introducing a large-scale *SAR* image database. Precisely, our dataset contains more than $60'000$ image instances and respective labels, chosen from 7 distinct semantic classes. Using this data, we perform a set of experiments to understand the impact of dataset size on classification accuracy. In this context, we also investigate the possibility to further expand the dataset with synthetic SAR images generated with the help of*Generative Adversarial Networks* (*GANs*). These are powerful generative models that have been shown to produce high-quality synthetic images in other fields, thereby reducing (or even compeltely avoiding) the annotation effort. Our main contributions in this work can be summarized as follow:

- We construct the first state-of-the-art CNN model pre-trained on large-scale SAR data.
- We investigate the possibility of transfer-learning from other pre-trained models based on optical images, and their impact on SAR image classification.
- We investigate the possibility of training also with artificial SAR data generated with a GAN.

2. RELATED WORK

In the field of SAR image analysis, the use of deep-learning methods, such as *CNNs*, is still in its infancy, mainly due to the limited availability of VHR data with asociated ground truth labels. We note that, in a detailed literature review, we did not find any work that relies on a large scale SAR-database to unlock the potential of deep neural networks. Moreover, there are no pre-trained networks for SAR images, which would facilitate the classification of SAR datasets for which there aren't enough training labels to learn a deep network from scratch.

Published work at the intersection of *SAR* imaging and deep learning are mainly focussed on *Target Classification*. Some representative works employ sparsely connected layers [1], limited training data [2] and domain-specific data augmentation methods [3]. In the field of *GAN*s for *SAR* data, some interesting results have been shown by [4], where authors constructed a generative deep model. The outcome of their experiments however remain unconclusive, due to the scarcity of training data, and particular characteristics of the underlying targets (military imagery). Another implementation of *GANs* in the field of Remote Sensing is the one of [5], who investigate the *Wasserstein GAN* for poverty mapping with sparse labels, using a semi-supervised approach. They however do not use SAR imagery. Yet another work

*Authors have contributed equally in this work on optical remote sensing imagery and artificial data generation is the one of [6]. Thery propose an additional objective function over the standard GAN architecture to improve the output. While the approach is interesting, it ultimately does not produce visually realistic images of the target classes. A promising work is [7], which demonstrates the generation of synthetic SAR images on the basis of optical images. The high-quality samples generated in that work show the potential of *GAN* methods for SAR image synthesis, and motivate us to further investigate that topic.

3. THE DATASET

Our dataset was obtained via a novel classification scheme especially designed for high-resolution SAR imagery of (mainly)q built-up areas. The dataset contain image patches from 288 TerraSAR-X image scenes (41 scenes acquired in Africa, 6 from Antarctica, 59 from Asia, 80 from Europe, 40 from the Middle East, 54 from North and South America and 8 from ocean surfaces), with a total of over $60'000$ individual patches. All *TerraSAR-X* data are obtained via the X-band instrument, using the high-resolution Spotlight mode. The incident angles throughout the scenes varies between 20 and 50 degrees. The resolution of the images scenes is set to 2.9m, with a pixel spacing of 1.25m. The chosen polarization for the dataset is horizontal (HH) for all products. Furthermore, for conveniennce we convert all intensity data to 8-bit integer precision. For more information on the dataset, refer to [8].

4. EXPERIMENTS

In our experiments, we first set a baseline for deep learning based SAR classification, and go on to investigate if we can improve over that baseline with additional, synthetic data generated with a *GAN*.

4.1. The CNN SAR classifier

To establish a baseline for the use of CNNs with SAR data, we employ a state-of-the-art network architecture for optical images, namely the standard Residual Network with 50 hidden layers (*ResNet-50*) [9]. To adapt the network to our class nomenclature, we remove the fully connected layers at the top and replace them with three fully connected layers of size $256, 256$ and 7, respectively, which we train from scratch. The resulting model achieves an overall accuracy of 93.2%. We find this result very encouraging: in spite of the radically different imaging process and image statistics, modern, deep CNNs appear to be suitable for supervised SAR image classification and yield high classification accuracy, when trained on an appropriate, large training set.

A further, interesting observation is that conventional pre-training (i.e., initialization with the weights learned from optical images) has little effect on the classification result. This is not unexpected – while the pre-training with very large databases (millions of images) does ususally help when working with optical images, the local image statistics of RGB and *SAR* data are probably too different to transfer even low-level image properties. To support that hypothesis, we have we trained the same ResNet-50 twice, once with random initialization and once with weights pre-trained on *ImageNet*. The classification results for SAR were practical the same in both cases. I.e., the pre-trained weights do not hurt the learning, but they also do not help compared to random initialisation.

4.2. Image Generation with BEGAN Models

Given the good performance of the deep network, and the still comparatively small training database (in computer vision, models are routinely pre-trained with more than 10^6 training images), we investigate if artificial data generation with a *GAN* can further improve our classifier. For close-range applications, it has already been shown that classifier training can benefit from *GAN* image synthesis, e.g., for sign recognition [10]. However, our task however is more challenging, due to the extreme variability of the SAR data in our database, and the large dimension of the output images we need to generate (160×160 pixels).

4.2.1. BEGAN Model for SAR

Despite the rather recent invention of *GANs*, there is already a plethora of variants such as *DC-GANs*, *cGANs*, *WGANs*, *DRAGANs* and *BEGANs*. We base our investigation on the newly proposed *BEGAN* model [11], which was shown to generate images of remarkable quality, and to handle larger image sizes than most other variants.

Compared to the standard *GAN* model, the *BEGAN* design has a number of attractive characteristics. First, it uses autoencoders as discriminator, thus matching the corresponding autoencoder distributions (rather than the rawe data distributions), with a Wasserstein distance loss. Furthermore, *BEGAN* employs an equilibrium term to balance the effect of the *Discriminator* with respect to the *Generator*, so as to avoid an "early win" of one stage over the other.

BEGAN was initially proposed for generating human faces. Even though this is already a challenging problem, synthesising SAR images proved to be a lot harder. Through empirical experimentation, we found that the capacity of the original model is not sufficient to capture the complexity of our database. We therefore added more layers both to the *Generator* and the *Discriminator*. In each of the two stages, we add two additional convolution layers (with respective eLU non-linearities), before the respective pooling/upsampling layers. Furthermore, we have replaced the final, linear layers of both stages with non-linear ones, using the ReLU non-linearity. [1]

[1] **Code**: `https://github.com/deep-unlearn/Big_Data_From_Space_2017`

Finally, and perhaps most significantly, we have changed the loss function of the discriminator.The original loss function is simply the mean of the per-pixel L_1 distance. In our model, we replace it by a combination of a per-pixel distance and a histogram distance, to explicitly match the global intensity distributions of the images. The new loss is given by :

$$\mathcal{L}_{\text{generated}} = L_{\text{hist}} + \omega \cdot L_{\text{spatial}}$$
$$L_{\text{hist}} = \frac{1}{N_{\text{bins}}} \cdot \sum(\text{hist}(X) - \text{hist}(X_{\text{recon}}))^2$$
$$L_{\text{spatial}} = \frac{1}{N_{\text{pix}}} \cdot \sum(X - X_{\text{recon}})^2 \ ,$$

where *hist* returns the histogram of an image over a fixed number N_{bins} of bins (set to 64), and N_{pix} is the number of pixels in the generated image. The hyperparameter ω defines a weighting between the two parts of the loss. For our experiments we empirically set it to $\omega = 0.001$.

4.2.2. *BEGAN Image Generation*

Image generation with GANs still remains somewhat a brittle and somewhat challenging task. We thus investigate three for our SAR image generation problem. They are:

- In the hard scenario, the network is asked to directly generate large SAR patches of size 160×160 pixels. This scenario would be optimal, in the sense that it outputs patches at the correct size for our database; but it also the most complex prediction task.
- In the intermediate scenario, the network generates SAR patches at $2\times$ larger GSD, with dimension 80×80 pixels, which are then compared to downsampled real images. The reduced resolution lowers the complexity of the task, while the patch size in scene coordinates, and thus the spatial context, remains the same. But the resulting images must be upsampled to the original dimensions, and thus lack high-frequency detail.
- In the simple scenario, images are also generated at 80×80 pixels, but this time the original GSD is retained. Instead, the patch size in scene coordinates is halved, respectively the real SAR patches are cropped. The resulting images must again be upampled, to match the patch size used for classification. Using smaller *and* more local patches presumably further reduces the complexity of the prediction, the price to pay is a mismatch in GSD between synthetic and real training images, and the loss of 3/4 of the context area.

So far, we were unsuccessful in our attempts to train the hard scenario. We leave it to future work to determine whether this can be remedied, or whether a higher-capacity model is needed. For the intermediate scenario, the generator appeared to converge better, but its outputs were still unsatisfactory and did not visually resemble the original data. For the time being, this failure leaves us with the simple scenario. That setting did converge to a reasonable solution that outputs realistically-lookinf synthetic images, see examples in *Figure 1* and real SAR data in *Figure 2*. However, one can also clearly see that the smaller patches capture less of the context.

4.3. Classification Augmentation Through GANs

In spite of the limited success to synthesize full-size patches, we continued the experiment. The “simple”patches were upsampled to 160×160 pixels and added to the training data for the classification network. As a first test, we generated 5100 synthetic instances of the *Settlement* class, which is the most frequent class in the dataset (25’000 real traning patches), and also the one with the strongest intra-class variation.

Somewhat surprisingly, retraining the *ResNet-50* classifier with the augmented dataset did not influence the classifier either way. We get the same classification accuracy of 93.2%. Seemingly, the synthetic examples were neither capable of adding any additional information that would have improved the classifier, nor were they unrealistic enough to negatively impact the classifier. Obviously, in the absence of a satisfactory explanation such an outcome appears unlikely. Future work will have to determine the cause, and hopefully address the current short-comings of the generator, so as to further improve the classifier network.

5. CONCLUSIONS

We have introduced a new, large-scale database of SAR patches with asociated semantic class labels. To our knowledge, this is the first SAR dataset large enough to train modern deep neural networks, and we have demonstrated that capability by learning a *ResNet-50* convolutional network that achieves an excellent 93.2% hit rate over 7 different scene categories. We have further adapted the generative BEGAN network model to SAR data, and have experimented with synthetically generated images to obtain an even larger traning set. Unfortunately, we are still struggling with technical problems in the image synthesis, and the first experiments with additional, synthetic training data have not yet led to conclusive results. Nevertheless, our paper clearly shows that, as soon as enough data is available, deep convolutional networks work extremely well also for SAR images. More detailed tests and comparisons still need to be run, but we believe that our results set a new standard for patch-wise SAR classification. We also posit that our failure to exploit synthetic images is due to relatively minor technical difficulties that can be addressed, and we are still convinced that GANs have the potential to support the the generation of truly big training databases.

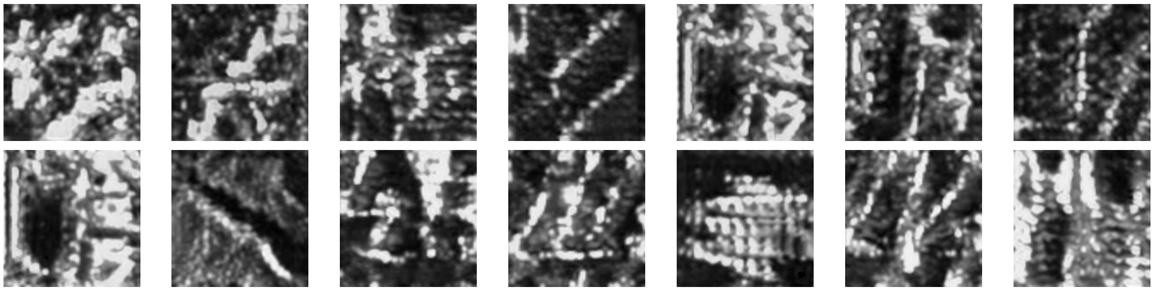

Figure 1. Generated data of size 80×80 pixel by cropping scenario - upsampled to 160×160 pixel

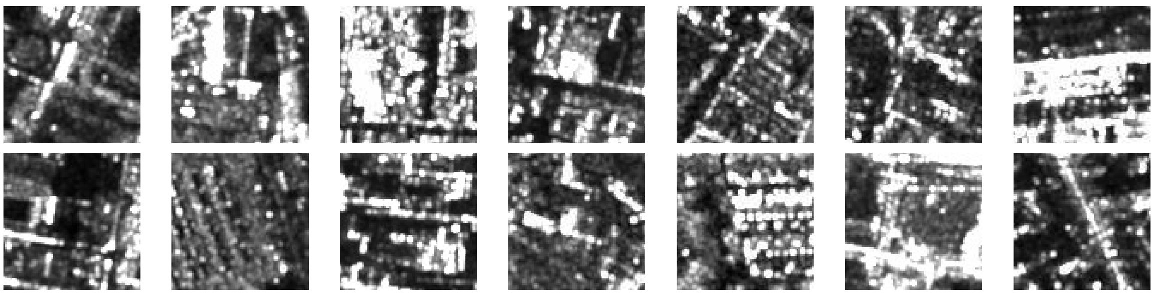

Figure 2. Original *TerraSAR-X* data of original size - 160×160 pixel

6. REFERENCES

[1] Sizhe Chen, Haipeng Wang, Feng Xu, and Ya-Qiu Jin, "Target classification using the deep convolutional networks for sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.

[2] Zhao Lin, Kefeng Ji, Miao Kang, Xiangguang Leng, and Huanxin Zou, "Deep convolutional highway unit network for sar target classification with limited labeled training data," *IEEE Geoscience and Remote Sensing Letters*, 2017.

[3] Jun Ding, Bo Chen, Hongwei Liu, and Mengyuan Huang, "Convolutional neural network with data augmentation for sar target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364–368, 2016.

[4] Jiayi Guo, Bin Lei, Chibiao Ding, and Yueting Zhang, "Synthetic aperture radar image synthesis by using generative adversarial nets," *IEEE Geoscience and Remote Sensing Letters*, 2017.

[5] Anthony Perez, Swetava Ganguli, Stefano Ermon, George Azzari, Marshall Burke, and David Lobell, "Semi-supervised multitask learning on multispectral satellite images using wasserstein generative adversarial networks (gans) for predicting poverty," *Technical Report*, 2017.

[6] DaoYu Lin, "Deep unsupervised representation learning for remote sensing images," *arXiv preprint arXiv:1612.08879*, 2016.

[7] Nina Merkle, Peter Fischer, Stefan Auer, and Rupert Müller, "On the possibility of conditional adversarial networks for sar template generation," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017.

[8] Corneliu Octavian Dumitru, Gottfried Schwarz, and Mihai Datcu, "Land cover semantic annotation derived from high-resolution sar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 6, pp. 2215–2232, 2016.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] Xinlong Wang, Mingyu You, and Chunhua Shen, "Adversarial generation of training examples for vehicle license plate recognition," *arXiv preprint arXiv:1707.03124*, 2017.

[11] David Berthelot, Tom Schumm, and Luke Metz, "Began: Boundary equilibrium generative adversarial networks," *arXiv preprint arXiv:1703.10717*, 2017.