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]Department Photogrammetry & Remote Sensing, 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 for exploring patch-based classification from 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 access if the derived data can actually improve classification accuracy, with no additional annotation.

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

1. INTRODUCTION

Classification of very high resolution (VHR) *SAR* image data remains still a hard and time-consuming task. Some of the major difficulties involve, the scarcity of available data, the challenging interpretation of the semantic content and the particular characteristics of the *SAR* derived scattered signal. All these important factors lead to a general absence of large scale *SAR-derived* image databases for Remote Sensing image analysis and knowledge discovery. Furthermore, despite the major advances on Big Data image classification, namely *Deep Learning* methods, SAR-based systems have shown to benefit little from these important breakthroughs, mainly due to the limited availability of data and/or respective labels.

In this work we tackle the problem of limited data by introducing and experimenting with a large-scale *SAR* derived image database. Precisely, our dataset contains more than 60000 image instances and respective labels, chosen from 7 distinct semantic classes. Using this large-scale dataset, we perform a set of experiments for understanding the impact of data size in relation to classification accuracy. In this context, we also investigate the possibility of enhancing our dataset by generating artificial SAR images through the use of *Generative Adversarial Networks* (*GANs*). Through this experiment, we access the possibility of artificial data generation for decreasing or completely avoiding the time-consuming data annotation process. Our main contributions in this work can be summarized as follow:

- We construct the first large-scale pre-trained SAR image model, based on deep-learning Convolutional Neural Networks learning paradigm.
- We investigate the possibility of transfer-learning from other pre-trained computer-vision based models and their impact on SAR image classification.
- We investigate the impact of artificially derived SAR data over the classification accuracy, by employing state-of-the art Generative Adversarial Networks (GANs).

2. RELATED WORK

In the field of SAR image analysis the use of deep-learning methods, such as *CNNs* models is still in its infancy stage, mainly due to limitations on availability of VHR data and respective labels. Importantly, despite our detailed investigation in literature we did not find any work employing large scale SAR-databases for alleviating the full potential of *CNN-based* methods. Furthermore, there are no available pre-trained, SAR-based models for facilitating the classification of small-size SAR image datasets.

Published work in the intersection of *SAR* imaging and deep-learning are mainly focusing on the topic of *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* model and *SAR* data some interesting results have been proposed by [4], where authors constructed a generative framework. The outcome of their experiments however had limited success due to the scarce training data and particular characteristics of the underlying targets (military imagery). Another

*Authors have contributed equally in this work implementation of *GANs* in the field of Remote Sensing is this of [5], where authors investigate the *Wasserstein GAN* for poverty mapping with sparse labels using a semi-supervised approach. In this work however despite the interesting findings, results are not based on SAR imagery. Yet another work on optical remote sensing imagery and artificial data generation is this of [6]. In this work the author proposes an additional objective function over the standard GAN architecture, for improving the generated data. Despite the interesting framework however, the derived images are of poor quality and do not resemble any true semantic class. A promising work is this of [7] where authors show high-quality image results on SAR generated images using as a base optical data. The samples generated here are of exceptional quality, showing the potential of *cGAN* methods in SAR image generation.

3. THE DATASET

Our dataset was obtained via a novel classification scheme especially designed for high-resolution radar SAR imagery, mainly depicting built-up areas. The dataset contain information from 288 TerraSAR-X image scenes (41 scenes are acquired from 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 over 60000 individual image patches. All *TerraSAR-X* data are obtained via the X-band instrument using the high-resolution Spotlight mode. The incident angle 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 set to horizontal model (HH) for all products. Furthermore as a standarization process, we converted all intensity data to 8-bit integer precision. For more information on the dataset one can refer to [8].

4. EXPERIMENTS

In our experiments, we initially set a strong baseline through a deep-learning classification scheme and further investigate if we can improve accuracy by introducing artificially *GAN* generated data.

4.1. The CNN SAR classifier

As a baseline for our experiments, we employ a state-of-the-art CNN classification architecture, namely the Residual-Network that contains 50 hidden layers (*ResNet-50*) [9]. For our purposes, we remove the fully connected part of the architecture and replace these layers with three fully connected layers of size $256, 256$ and 7 respectively. With this particular model, we achieve an overall classification accuracy of 93.2%, proving the strong properties of CNN networks.

Yet another interesting observation is related with the fact that pre-trained networks (optimum weights initialization) does not seem to affect the overall result. This may be a logical outcome, considering that pre-trained networks are generally trained on standard RGB images that bear very different properties with respect to *SAR* images. Under this observation, would be therefore normal to assume that pre-trained weights do not to have any significant effect over the training process. For proving this hypothesis, we trained our CNN network both with random initialization weights and *ImageNet* pre-trained weights. The results in both case were similar with little to no-variation.

4.2. Image Generation with BEGAN Models

In this section, we investigate if artificial data generation through *GAN* models can actually improve our classification baseline. Theoretically, such a task should be feasible considering previous success is literature in tasks such as sign recognition [10]. Our task however is much more challenging, considering the extreme variance of our SAR dataset and the larger image dimensions required to be generated (160×160 pixels).

4.2.1. BEGAN Model Selection

In the context of adversarial networks (*GANs*), there is a plethora of model variations such as *DC-GANs*, *cGANs*, *WGANs*, *DRAGANs* and *BEGANs*. For our model, we decided to investigate the newly proposed *BEGAN* model [11], as it has shown remarkable quality of generated image results, in addition to large image sizes.

BEGAN model has some advanced construction characteristics in relation to standard *GAN* models. Precisely, some of this unique characteristics are related with the use of autoencoders as a discriminators, therefore instead of directly matching data distributions, *BEGAN* matches autoencoder distributions through a Wasserstein loss distance measure. Furthermore, *BEGAN* employs an equilibrium term that tries to balance the effect of the *Discriminator* in respect to the *Generator* so there is no 'early' win of one model over the other.

4.2.2. BEGAN Model Modifications

BEGAN model was initially proposed for generating human faces. Even though this is considered a challenging problem, modeling SAR images has proven a much harder task. Through empirical experimentation, we found that the capacity of the original model will not suffice to model all of our data complexity, hence assigned additional processing layers both to the *Generator* and *Discriminator* parts of the network. Precisely, we have enhanced the network by adding two additional convolutional layers (with respective eLU non-linearities) on every level, before respective pooling/ upsampling layers. Furthermore, we have replaced the last linear layers of both the *Generator* and the *Discriminator*, with non-linear ones (ReLU non-linearity) [1].

The most significant modification is the introduction of a new loss measure within the discriminator model. In detail, the original loss-metric is a simple per-pixel L_1 mean difference. In our model, we have replaced this simple metric with a combination of spatial and pixel-value losses , for better modeling the quality of generated images. Precisely, the new image metric is given by :

$$\mathcal{L}_{geneated} = L_{hist} + \omega \cdot L_{spatial}$$

$$L_{hist} = \frac{1}{N_{bins}} \cdot \sum (hist(X) - hist(X_{recon}))^2$$

$$L_{spatial} = \frac{1}{N_{pix}} \cdot \sum (X - X_{recon})^2$$

Where, *hist* provides the histogram of values for a fixed number of bins (set empirically to 64), N_{bins} is equal to the number of histogram bins used and N_{pix} is equal to the number of pixels in the generated image. Furthermore hyperparameter ω defines a weighting for the spatial loss-metric. For our experiments we found that a value of $\omega = 0.001$ works optimally in our case. Importantly, the $\mathcal{L}_{generated}$ in our model relates to $\mathcal{L}_{\mathcal{D}}$ in the original *BEGAN* model. However our notation seems more proper as it underlines its reference to the generated image-loss.

4.2.3. BEGAN Image Generation

Image generation though GAN still remains an extremely challenging and complex task to achieve. For this reason, we investigated three independent scenarios for our SAR image-generation problem. These scenarios are:

- The hard scenario investigates the direct generation of large SAR images of size 160×160 pixel. This scenario is the be optimal as it produces images with the standard input dimensions, hence no further processing is required. Is important however to mention that this scenario contains the higher model complexity.

- The intermediate scenario investigates the possibility of downsampled SAR image generation of size 80×80 pixel. This scenario allows a simpler version of the data to be modeled due to downsampled information, however the data still retain the complete contextual information which is crucial for the semantic classification. Generated data with this paradigm require an upsampling to match the original spatial image dimensions.

- The simple scenario investigates the possibility of modeling images of size 80×80 pixels, produced by cropping the original data. This option affects the contextual information of the data but preserves the original ground sampling distance. In theory this reduced information would be easier to model due to decreased information complexity. Furthermore, these data need to be upsampled to be used along with our image classifier. This upsampling however does not grantees a good result as 3/4 of the original information is lost (cropping).

Despite our effort to directly produce complete image data (hard-scenario) this was proven infeasible. It seems that the complexity of the scenes is too great to be modeled adequately with our proposed modified *BEGAN* model. This failure also resulted for the downsampled data of the intermediate-scenario. In this case the generator seems to converge to a better solution, however the optical result were far from optimal, with little relation to the original data. Finally the simple-scenario of cropping 80×80 pixel from the original data was the one that produced good overall results. Generated image can be seen in image *Figure 1* and compared to the original data of *Fifure 2*. Nevertheless, the decreased spatial context is clearly visible as in this case we modeled a limited extent of the original data.

4.3. Classification Augmentation Through GANs

Despite our limited success to generate full contextual image patches, we decided to further investigate our generated data by upsampling and incorporating them to our original training data. For this purpose we generated 5100 new training instances of the class *Settlements* using the *simple-scenario* described before. Importantly the *Settlements* class is our most complex and most frequent semantic class, with over 25000 instances in our original classification dataset. Therefore, we decided to uniquely focus on this class for our experiments in this work.

For this experiment we retrained our standard *ResNet-50* classifier including the new data and achieved an unexpected outcome. The overall accuracy was again 93.2%, exactly as before the data augmentation. This result was quite unexpected as it did not diminish or improved the classification.

5. CONCLUSIONS

In this work, we introduced a new large-scale SAR database and achieve great performance on classifying it in seven distinct semantic classes. We further generated artificial data using the BEGAN adversarial paradigm with some restrictions regarding the depicted context. In our future work work, we plan to expand this line of research by exploring conditional-GAN models for modeling the complete semantic context.

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

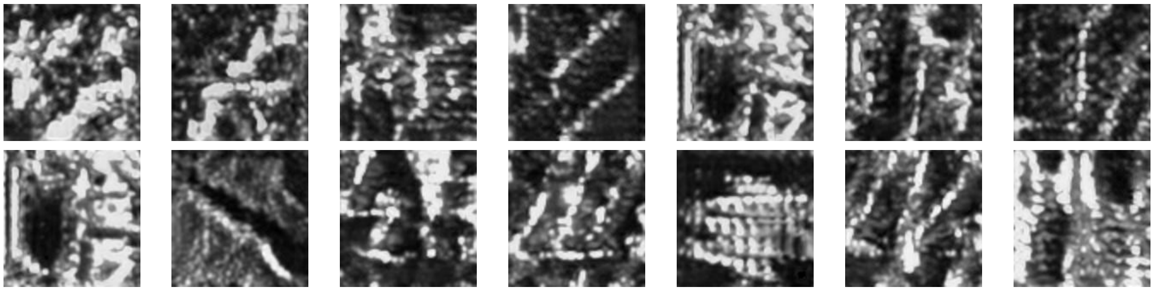

Figure 1. Generated data of size 80×80 pixel by cropping scenario - upsampled to 160×160 pixel

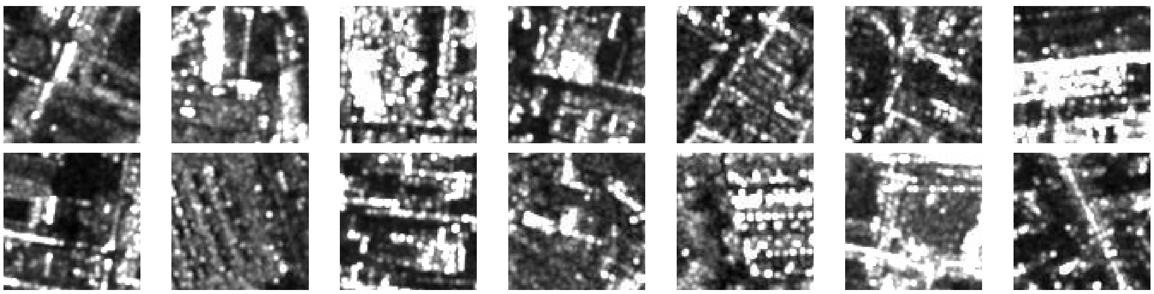

Figure 2. Original *TerraSAR-X* data of original size - 160×160 pixel

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.