ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Detection and Localization of GAN Manipulated Multi-spectral Satellite Images

Lydia Abady, Giovanna Maria Dimitri and Mauro Barni *

University of Siena - Department of Information Engineering and Mathematics
Siena - Italy

Abstract. Owing to their realistic features and continuous improvements, images manipulated by Generative Adversarial Network (GAN) have become a compelling research topic. In this paper, we apply detection and localization to GAN manipulated images by means of models, based on EfficientNet-B4 architectures. Detection is tested on multiple generated multi-spectral datasets from several world regions and different GAN architectures, whereas localization is tested on an inpainted images dataset of sizes 2048×2048×13. The results obtained for both detection and localization are shown to be promising.

1 Introduction

In recent years, the development of deep learning (DL) techniques for image forgery and authenticity verification has seen a steep increase in terms of research interest [1]. The importance of the application of such techniques to satellite images is manifold, for instance to fend off misinformation campaigns. A few examples are present in the literature. For instance, in [2] the authors propose a two steps workflow for forgery detection and localization in satellite imagery. The first step makes use of a generative adversarial network (GAN) in order to obtain features representation of the pristine satellite images. This first step is followed by a one-class SVM classifier. In [3], instead, the authors use a conditional GAN architecture trained using two classes, where the generator is employed to estimate the forged mask. Another notable example is [4], where the authors use an approach similar to [2] with the difference of training jointly the autoencoder and a SVDD (Support Vector Data Descriptor). While in [5] the authors proposed the use of a GAN-based inpainting nested U-net, trained to estimate heatmaps of forged images. To the best of our knowledge, however, all of the works present in the literature tackle the research problem using solely 3 bands images.

In our work, we propose to use either the full 13 bands of Sentinel-2 level1-C samples or 4 bands out of the 13 bands for image forgery localization and detection, obtaining promising results. The forged datasets are either a translation of

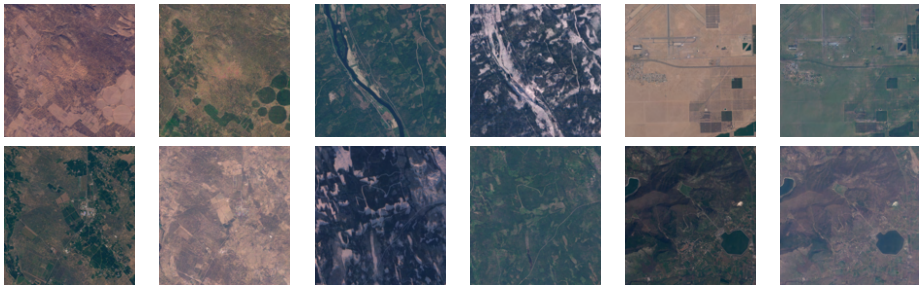
*This material is based on research sponsored by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) under agreement number FA8750-20-2-1004. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or AFRL or the U.S. Government

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

an image from barren to vegetation and vice versa or a translation from winter to summer and vice versa. The paper is structured as follows. In Section 2 we present an overview of the datasets used. In Section 3 the methodology employed is sketched with the respective results and in Section 4 we draw conclusions and outline future perspectives.

2 Datasets

All of the images used in the experiments are of two types: pristine Sentinel-2 level1-C, or images generated from GAN architectures trained on Sentinel-2 level1-C datasets. We downloaded the pristine images from the ESA Copernicus hub [6]. The images are made up of 13 bands, four of which, bands 2 (Green), 3 (Blue), 4 (Red) and 8 (Near InfraRed NIR), have a spatial resolution of 10m and a pixel resolution of 10980×10980. Six bands have a spatial resolution of 20m. The remaining 3 bands have a spatial resolution of 60m. All bands have a radiometric resolution of 16 bits. For the 13 bands datasets, we up–sampled all of the bands not having a spatial resolution of 10m, in order to have the same resolution of 10980×10980 analogously to the 10m bands. Subsequently the images with all 13 bands were tiled into 512×512 resolution patches by using gdal-retile from the gdal software library [7]. For what concerns the 4 bands datasets, instead, we extracted only the 10m bands and then re–tiled them into 512×512 pixel resolution. Table 1 shows a summary of all the listed datasets below.



(a) China Pristine | (b) China Generated (season transfer) | (c) Scand Pristine | (d) Scand Generated (season transfer) | (e) LC Pristine | (f) LC Generated (landcover transfer)

Fig. 1: RGB examples of the Datasets.

2.1 Land cover transfer datasets

In an effort to create these land cover datasets, we trained two cycleGAN [8] models, one for the 13 bands samples and another for the 4 bands samples. The models were trained to transfer barren to vegetation landscapes and vice versa. The training datasets were obtained having in mind that one domain should contain mainly images dominated by vegetation while the other by barren terrain. For that purpose, utilizing the organization for economic co-operation and

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

development land cover classifications [9], we downloaded images from Salvador, Congo, Montenegro, Gabon and Guyana, for what concerns the vegetation domain, while we downloaded images from South and Central America, for the barren domain. Subsequently a total of 8000 images of 512×512 resolution per domain were used to train the GAN models. Furthermore, we used 2000 images per domain to obtain 2000 style transferred images per domain. We therefore obtained two datasets: one from the models trained on the 4 bands datasets, and a second one from the models trained on the 13 bands datasets. Each dataset contains 8000 images 4000 of which are pristine (2000 barren and 2000 vegetation) and 4000 are generated (2000 barren and 2000 vegetation). In Figure 1e, we show a couple of examples of the pristine images while Figure 1f shows a couple of examples of the generated images.

2.2 China season transfer datasets

To build China dataset, we trained 2 pix2pix models [10] in order to transfer summer images into to winter ones and vice versa. For each transfer direction, we had to train a model, one that learns how to transfer from summer to winter and another that learns how to transfer from winter to summer. In addition to this, we had to train 2 models for the 13 bands datasets and 2 models for the 4 bands datasets. Training the models requires a paired dataset, therefore we downloaded images located in China from the same exact region, but referring to two different months. In particular images downloaded from the month of August 2020 refer to the summer domain, while those downloaded in January 2021 belong to the winter domain. Overall the models were trained on 6000 coupled images. Consequently, we obtained two datasets, one from the models trained on 13 bands and the other from the models trained on 4 bands. Each dataset contains 8000 images where 4000 are pristine images (2000 summer and 2000 winter) and 4000 generated images (2000 summer and 2000 winter). In Figure 1a, we show a couple of examples of the pristine images, while Figure 1b shows two examples of the generated images.

2.3 Scandinavian season transfer datasets

We built the Scandinavian season transfer datasets similarly to the China season transfer dataset, by only changing the region from China to Scandinavian countries. Images that represent the summer domain were downloaded in June 2020, whereas for the winter domain, the images correspond to February 2020. In Figure 1c, we show a couple of examples of the pristine images, while Figure 1d shows a couple of examples of the generated images.

2.4 Inpainted dataset

This dataset was built to test localization of the manipulated content. The size of the images is 2048×2048 for all the 13 bands. As a first step, we tiled images gathered for the vegetation domain, as described in Section 2.1. Afterwards we

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

used the vegetation to barren GAN model on a selected region of each image. To pick the to-be-manipulated region, we segmented a gray scale version of the image in order to obtain the binary mask with the region to be in-painted. After that, we detected the contours of the binary mask in order to locate the largest contour within a bounding box of 512×512. Furthermore we replaced the region within the contour with the transferred image after applying a Gaussian blur to the boundaries of the contour. Acting in this way we obtained a total of 50 manipulated images.

	Channels	Resolution	Architecture	Type of Transfer	Number of Images
Land Cover Dataset (LC)	13 bands or 4 bands	512x512	CycleGAN	Land Cover	8000
Scandinavian Dataset (Scand)	13 bands or 4 bands	512x512	Pix2pix	Season	8000
China Dataset	13 bands or 4 bands	512x512	Pix2pix	Season	8000
Inpainted dataset	13 bands	2048x2048	–	–	50

Table 1: List of generated datasets

3 Methodology and Evaluation

In order to detect if the images are fully manipulated or pristine, we trained the base EfficientNet-B4 network (eff) with similar settings as the original paper [11], with the only exception of varying the input size to fit our datasets (13 bands or 4 bands). Training was carried out on either the LC dataset, the Scand dataset or both datasets combined, then we cross– tested those models on all the generated datasets. As proposed in [12], to enhance generalization, we also trained EfficientNet-B4 with no down (eff_nodown) where we refrained from applying down-sampling in the initial layer. We compared the results with the base EfficientNet-B4 models. All the models were trained with augmentation, by applying Gaussian blur, random shift, random rotation and random flip. In Table 2, we show the cross testing results of detection accuracy, on the 13 bands datasets. The results show good detection performances when the datasets are matched between training and testing (i.e. generated with the same GAN architecture). Also, the results we got might suggest that the eff_nodown architecture has better generalization capabilities than the base architecture. In addition, in Table 3, we show the results of the models trained on the datasets generated by GANs using the 4 bands input and tested on datasets generated by GANs with 4 bands input and 13 bands input where we extracted only the 4 10m sampled bands. The results show that the models generalize pretty well on the cross datasets except for the LC dataset, for which detection only works in the case where the training and the testing datasets are generated by the same GAN.

With regard to *localization*, we trained several models similar to the eff_nodown, on both LC and Scand datasets combined, but with varying input sizes of 32×32, 64×64, 128×128, 256×256 and 512×512 (the various image sizes were obtained by cropping the original 512×512 images datasets). Then, using the trained models, we applied a sliding window with stride 8, for the 32 input size and 64 input size models and stride 20 for the rest on the inpainted dataset. For each

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Results of 13 bands		TEST							
		LC		Scand		LC and Scand		China	
		eff_down	eff_nodown	eff_down	eff_nodown	eff_down	eff_nodown	eff_down	eff_nodown
TRAIN	LC	94.5	**99.75**	52.85	96	75.2	97.66	53.6	64.92
	Scand	50.0	50	**100**	99	75.27	74.2	52.41	58.2
	LC and Scand	93.85	97.25	99.85	99.9	96.85	**98.57**	50.24	**72.56**

Table 2: EfficientNet detection results base vs. nodown of 13 bands

Results of 4 bands		TEST					
		LC		Scand		China	
		4 bands	4 out of 13	4 bands	4 out of 13	4 bands	4 out of 13
TRAIN	LC	100	**67.5**	84	99.65	90.62	99.6
	Scand	54.75	50.15	100	96.7	93.1	70.42
	LC and Scand	**100**	61.3	**100**	**100**	**98**	**99.65**

Table 3: EfficientNet detection results of 4 bands

prediction, only the centered pixels of the stride×stride are classified. For the hierarchical models, we first used the 512×512 model with a stride of 100 to obtain the mask. Afterwards, we multiplied the obtained mask with the image and then we applied a smaller model (64 or 32) with stride 12 on the non-zero pixels. Table 4, shows the Dice coefficient and the Jaccard index for all the window sizes. The best results are achieved when using the hierarchical setup of 512 and 64. Figure 2 shows an example of the obtained mask in each setting.

	512	256	128	Hierarchical (512, 64)	Hierarchical (512, 32)
Jaccard Index	0.68	0.77	0.81	**0.84**	0.8
Dice Coefficient	0.77	0.84	0.88	**0.9**	0.87
Recall	**0.94**	0.91	0.91	0.86	0.83

Table 4: Localization scores using EfficientNet models

(a) Image (b) Ground Truth (c) 512 model (d) 256 model (e) 128 model (f) 512-64 model (g) 512-32 model

Fig. 2: Localization Examples

4 Conclusion

In this work, we presented a novel application of EfficientNet–B4 architectures for the detection and localization of multi-spectral manipulated GAN images. We applied the EfficientNet–B4 architectures to detect manipulated satellite images, in particular to 13 or 4 bands of Sentinel-2 level-1C manipulated datasets, generated by cycleGAN and pix2pix architectures. The manipulations were either global or local. The results we obtained show that in a matched detection

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

scenario (where the training and test sets are generated by the same GAN architecture), we are able to achieve a detection accuracy above 98%, while in the unmatched scenario improvements are still needed. In the case of localization, the best Jaccard index and Dice coefficient we obtained were 0.84 and 0.9 respectively. Future extensions of the present work concern the improvement of the generalization of detection, by using a one-class classifier and the creation of a more challenging inpainted dataset to test localization.

References

[1] Pengpeng Yang, Daniele Baracchi, Rongrong Ni, Yao Zhao, Fabrizio Argenti, and Alessandro Piva. A survey of deep learning-based source image forensics. *Journal of Imaging*, 6(3):9, 2020.

[2] Sri Yarlagadda, David Güera, Paolo Bestagini, Fengqing Zhu, Stefano Tubaro, and Edward Delp. Satellite image forgery detection and localization using GAN and one-class classifier. In *Electronic Imaging (EI)*, 2018.

[3] E. R. Bartusiak, S. K. Yarlagadda, D. Güera, P. Bestagini, S. Tubaro, F. M. Zhu, and E. J. Delp. Splicing detection and localization in satellite imagery using conditional GANs. In *IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2019.

[4] J. Horvàth, D. Güera, S. K. Yarlagadda, P. Bestagini, F. M. Zhu, S. Tubaro, and E. J. Delp. Anomaly-based manipulation detection in satellite images. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, 2019.

[5] János Horváth, Daniel Mas Montserrat, Edward J. Delp, and János Horváth. Nested attention u-net: A splicing detection method for satellite images. In Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, Giovanni Maria Farinella, Tao Mei, Marco Bertini, Hugo Jair Escalante, and Roberto Vezzani, editors, *Pattern Recognition. ICPR International Workshops and Challenges*, pages 516–529, Cham, 2021. Springer International Publishing.

[6] Copernicus open access hub. `https://scihub.copernicus.eu/dhus/#/home`. Accessed: 2019-2021.

[7] Gdal. `https://gdal.org/`. Accessed: September 2020.

[8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 2242–2251. IEEE Computer Society, 2017.

[9] Oecd. `https://stats.oecd.o/Index.aspx`. Accessed: September 2020.

[10] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[11] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 2019.

[12] Diego Gragnaniello, Davide Cozzolino, Francesco Marra, Giovanni Poggi, and Luisa Verdoliva. Are GAN generated images easy to detect? A critical analysis of the state-of-the-art. In *2021 IEEE International Conference on Multimedia and Expo, ICME 2021, Shenzhen, China, July 5-9, 2021*, pages 1–6. IEEE, 2021.