Investigating Model Robustness Against Sensor Variation

Matthieu Terris, Sagar Verma

INVESTIGATING MODEL ROBUSTNESS AGAINST SENSOR VARIATION

Matthieu Terris[1] and Sagar Verma[1,2]

[1] Granular AI, MA, USA
[2] CentraleSupélec, Université Paris-Saclay, France
matthieu.terris@gmail.com
sagar@granular.ai

ABSTRACT

Large datasets of geospatial satellite images are available online, exhibiting significant variations in both image quality and content. These variations in image quality stem from the image processing pipeline and image acquisition settings, resulting in subtle differences within datasets of images acquired with the same satellites. Recent progress in the field of image processing have considerably enhanced capabilities in noise and artifacts removal, as well as image super-resolution. Consequently, this opens up possibilities for homogenizing geospatial image datasets by reducing the intra-dataset variations in image quality. In this work, we show that conventional image detection and segmentation neural networks trained on geospatial data are robust neither to noise and artefact removal preprocessing, nor to mild resolution variations.

1. INTRODUCTION

Satellite data and geospatial machine learning offer an unparalleled source for objective global-scale data. The past decade has seen a nearly eight-fold increase in the number of earth observation satellites deployed to orbit. This increase represents a paradigm shift in the availability of satellite data and, thus, its potential in downstream applications. With imagery now available for nearly every place on earth daily, researchers can now rely on satellites for time-series data on natural and man-made changes. The ability to model earth's land use and land cover (LULC) and how it is changing due to human activities and natural phenomena will have manifold implications on the study of climate change, economic development, and anthropology. Further, it would serve as a valuable tool to improve decision-making in humanitarian aid and disaster relief efforts [1]. It would also provide critical insights into various topics such as sustainable development and urban sprawl, water and air contamination levels, and illegal construction.

The currently available geospatial imaging dataset shows great extrinsic and intrinsic diversity. As for extrinsic differences, BigEarthNet [2] images are created from Sentinel2 data, with a resolution of approximately 10m/pixel, while xView [3] has a 30cm/pixel resolution. Regarding intrinsic diversity, image data can show a strong signature given the satellite's orientation, camera sensitivity, altitude of the satellite, and image processing pipeline. Consequently, training models on satellite image data are prone to generalization issues on data acquired in a slightly different context [4].

Image processing techniques thus seem necessary for homogenizing the dataset, but the influence of the processing pipeline on the final segmentation or detection quality is unpredictable. For instance, it has already been observed that super-resolving geospatial images strongly hurt the quality of detection models [5, 6, 7]. Similarly, the variability of the image quality within a given dataset is acknowledged, and some works have been using restoration models to improve their quality [8]. However, little is known about the influence on the resulting segmentation/detection performance. In this paper, we propose to investigate state-of-the-art models for image restoration and image super-resolution [9, 10] and their influence on cornerstone tasks of geospatial imaging analysis, namely image detection and image segmentation.

2. PROPOSED APPROACH

Geospatial image datasets can contain both variations in noise and image processing artefacts [11] as well as mild variations in resolution. Given networks trained for image detection and segmentaion on traditional geospatial image datasets, we propose to investigate the impact of two preprocessing pipelines: artifacts (and noise) removal and super-resolution. The former allows generating images with the same high quality fixed for each image, while the latter allows fixing the per-pixel resolution of each image during training. We next briefly detail these two pipelines.

Recent works in blind image restoration (i.e. when no assumption is made on the type of image degradation, such as the nature of the noise or of the compression artefacts) have shown impressive results for natural images. Such setup applies to geospatial image datasets, where the image acquisition pipeline is often only partially known. We use the SCUNet neural network [10] as our artifacts removal pipeline, an architecture that has proven to be efficient for real image restoration tasks.

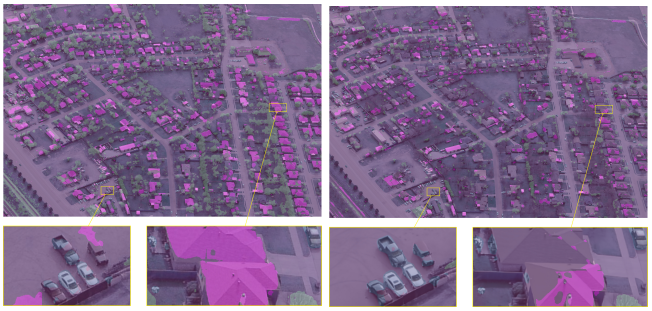
A longstanding limitation of super-resolution with deep neural networks was the restriction to integer upsampling factors. Recent works and new architectures have overcome this bottleneck [12, 13], allowing the proposal of meaningful super-resolution factors adapted to each image. In this work, we use the LTE super resolution network [13].

We use the YOLOv5 network [14] for both image detection and image segmentation; this network is trained on non-preprocessed versions of the datasets of interest.

3. EXPERIMENTAL RESULTS

3.1. Image segmentation after restoration

We show the influence of the noise and artifacts preprocessing with SCUNet neural network for image segmentation task in Figure 1. We notice that overall, the preprocessing tends to degrade the performance of the network. The roof class is less well detected (notice however that false positives also disappear), while the masks of cars are better recovered after the preprocessing step. This experiment suggests that the segmentation results strongly rely on statistical features, such as noise and artifacts that may not be visible to the naked eye. We however underline that preprocessing pipelines may be prone to errors, and in the particular example of Figure 1, the image restoration network over-smooths the image to the point of removing some visible tiles from the roof and replacing it with a uniform surface.



(a) No preprocessing (b) Noise and artefacts removal

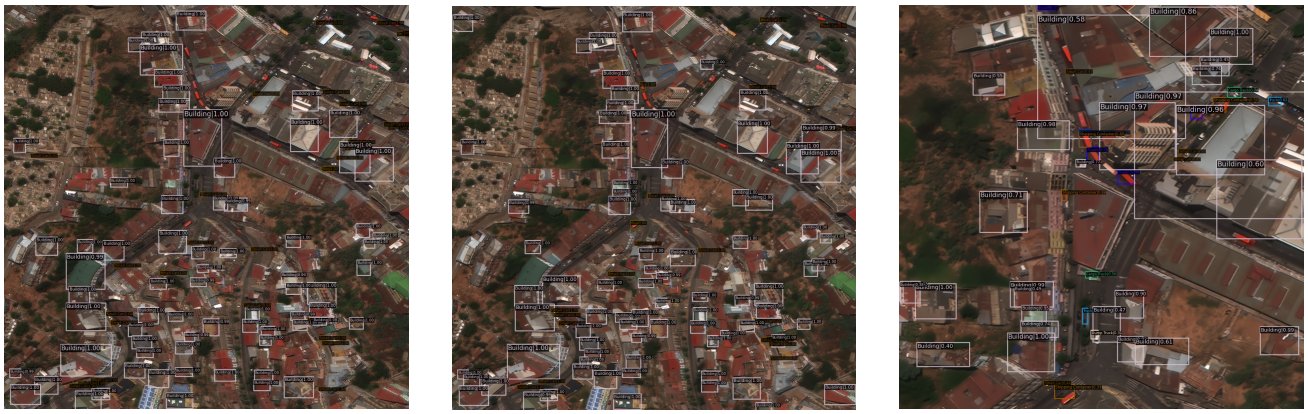
Fig. 1: Influence of noise and artifacts removal on the segmentation results. (a) shows segmentation masks when no preprocessing is performed on the input image; (a) shows segmentation masks when noise and artifacts removal is performed on the input image.

3.2. Image super resolution pipelines

We next propose to briefly investigate the performance of super-resolution methods for geospatial images on the BigEarthNet [2], DOTA[15] and xView [3] datasets. Since no groundtruth is available, we generate downsampled images with bicubic interpolation that will serve as our observations, with factors 1.5, 2 and 3. We next apply different super-resolution pipelines (namely bicubic interpolation, and LTE with SwinIR and RDN backbones) for these downsampled images and compute the PSNR between the recovered image and the original (groundtruth) image. Because geospatial images may suffer from strong artefacts, we also propose to add an additional artifacts removal step before performing super resolution. Metrics are given in Table 1. Overall, LTE with the RDN backbone performs best, with noticeable gains when an artifacts removal step is added prior to the SR network for DOTAv2.0 and BigEarthNet-S2. We notice that non-preprocessed LTE tend to perform poorly on the low quality image dataset BigEarthNet-S2, where bicubic interpolation works best.

3.3. Image detection after preprocessing

We next investigate the impact of variations in the resolution of the geospatial images on the image detection



(a) No preprocessing (b) Noise and artefacts removal (c) Super-resolved

Fig. 2: Experimental evidence of the influence of image restoration tools on the quality of Yolov5 object detection on the xView dataset. (a) shows results of image detection for a non-preprocessed image. (b) shows the result of Yolov5 on the same image, but that was preprocessed with SCUnet [10]. (c) Shows the results of Yolov5 on the same image, but super-resolved with a factor 1.6 with LTE [13]. Notice the different bounding boxes detected in the three cases.

Method	BigEarthNet-S2			DOTAv2.0			xView		
	×1.5	×2	×3	×1.5	×2	×3	×1.5	×2	×3
Bicubic	25.3	23.3	20.4	28.7	26.0	25.4	35.5	33.4	29.5
LTE-SWINIR	26.4	20.2	17.4	28.4	28.2	26.1	36.3	38.0	32.1
LTE-RDN	26.4	22.8	17.5	29.3	28.3	25.9	37.5	38.1	34.4
SCUNet + Bicubic	25.0	23.3	20.5	28.0	25.7	25.4	33.7	32.4	29.2
SCUNet + LTE-SWINIR	26.3	20.9	19.8	28.2	28.2	26.4	34.4	35.1	31.9
SCUNet + LTE-RDN	26.3	23.5	19.8	29.1	28.1	26.7	35.1	35.1	34.0

Table 1: Comparison of pure SR (top) and preprocessed SR (bottom). Best results are indicated in red, second best in blue.

pipeline. Visual results on the xView dataset are presented in Figure 2, and metrics on the full dataset are shown in Table 2. As a first observation, one notices that the building detections in the case of non-preprocessed and preprocessed images are fairly similar. On the opposite, a clear loss in performance for the building class is visible when processing a super-resolved image.

Metrics on the xView dataset are shown in Table 2. We notice a clear loss in performance when performing either preprocessing or super resolution with factor 1.6. Notice that the decrease in performance varies among classes; for instance, the "Car" and "Aircraft" classes are rather robust to preprocessing. Similarly, the resolution variation does not impact equally the performance among classes.

Further experiments have been performed on following datasets DOTA[15], Houston UAV[16], OSCD

Class	**Recall**			**AP**		
	w/o pre.	w pre.	SR	w/o pre.	w pre.	SR
Aircraft	0.263	0.285	0.208	0.212	0.208	0.145
Car	0.183	0.179	0.109	0.152	0.148	0.068
Truck	0.073	0.081	0.041	0.020	0.019	0.009
Train	0.186	0.133	0.070	0.109	0.073	0.029
Ship/Boat	0.319	0.272	0.197	0.163	0.139	0.069
Eng. Vehicle	0.144	0.126	0.048	0.068	0.060	0.016
Building	0.127	0.112	0.088	0.085	0.067	0.054
Mean (all classes)	-	-	-	0.096	0.082	0.044

Table 2: Metrics on the xView validation set in the three cases. Notice the substantial decrease of the mean AP metric as one progressively moves from no preprocessing (first and fourth column) to noise and artefact removal (second and fifth) to super-resolved (third and sixth). We report top-level hierarchy by grouping semantically similar objects to avoid clutter.

BiDate[17], OSCD MultiDate[18], SeeDroneSeaV2[19] and QFabric[20]. These datasets and experiments' outputs are available in the GeoEngine platform [21, 22].

4. CONCLUSION

In this paper, we have shown that applying state-of-the-art restoration models for artifact removal and image super-resolution within the imaging pipeline can strongly perturb the predictions despite a very mild influence on the naked eye. While image restoration may improve some of the segmentation masks for some spe-

cific classes, it leads to a drop in performance overall. Preprocessing geospatial image datasets with image enhancement networks may reduce the dependency of neural networks for downstream tasks to spurious noise and artifacts.

5. REFERENCES

[1] S. Verma, S. Gupta, and K. Gupta, "Aligning Geospatial AI for Disaster Relief with The Sphere Handbook," 2022.

[2] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "BigEarthNet: A large-scale benchmark archive for remote sensing image understanding," *IGARSS*, pp. 5901–5904, 2019.

[3] D. Lam, R. Kuzma, K. McGee, et al., "xView: Objects in context in overhead imagery," *arXiv:1802.07856*, 2018.

[4] M. T. Razzak, G. Mateo-García, G. Lecuyer, et al., "Multi-spectral multi-image super-resolution of sentinel-2 with radiometric consistency losses and its effect on building delineation," *ISPRS JPRS*, vol. 195, pp. 1–13, 2023.

[5] A. Van Etten, "You only look twice: Rapid multi-scale object detection in satellite imagery," *arXiv:1805.09512*, 2018.

[6] J. Shermeyer and A. Van Etten, "The effects of super-resolution on object detection performance in satellite imagery," in *CVPRW*, 2019.

[7] Y. Wang, S. M. A. Bashir, M. Khan, et al., "Remote sensing image super-resolution and object detection: Benchmark and state of the art," *ESA*, 2022.

[8] B. Rasti, Y. Chang, E. Dalsasso, et al., "Image restoration for remote sensing: Overview and toolbox," *IEEE GRS Magazine*, vol. 10, no. 2, pp. 201–230, 2021.

[9] J. Liang, J. Cao, G. Sun, et al., "Swinir: Image restoration using swin transformer," in *CVPR*, 2021, pp. 1833–1844.

[10] K. Zhang, Y. Li, J. Liang, et al., "Practical blind denoising via swin-conv-unet and data synthesis," *arXiv:2203.13278*, 2022.

[11] S. Saunier, F. Done, and C. Albinet, "Technical note on quality assessment for blacksky," *Eur. Space Agency, Paris, France, Tech. Rep., EDAP. REP*, vol. 26, 2021.

[12] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *CVPR*, 2021, pp. 8628–8638.

[13] J. Lee and K. H. Jin, "Local texture estimator for implicit representation function," in *CVPR*, 2022, pp. 1929–1938.

[14] G. Jocher and al., "ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements," 2020.

[15] G.-S. Xia, X. Bai, J. Ding, et al., "DOTA: A large-scale dataset for object detection in aerial images," in *CVPR*, June 2018.

[16] S. Goswami, S. Verma, K. Gupta, and S. Gupta, "FloodNet-to-FloodGAN : Generating Flood Scenes in Aerial Images," 2022.

[17] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *IGARSS*, 2018.

[18] M. Papadomanolaki, S. Verma, M. Vakalopoulou, et al., "Detecting urban changes with recurrent neural networks from multitemporal sentinel-2 data," in *IGARSS*, 2019, pp. 214–217.

[19] B. Kiefer, M. Kristan, J. Perš, et al., "1st workshop on maritime computer vision (macvi) 2023: Challenge results," in *WACV Workshops*, January 2023, pp. 265–302.

[20] S. Verma, A. Panigrahi, and S. Gupta, "QFabric: Multi-task change detection dataset," in *CVPRW*, 2021, pp. 1052–1061.

[21] H. Shin, N. Exe, U. Dutta, et al., "Europa: Increasing accessibility of geospatial datasets," in *IGARSS*, 2022.

[22] S. Verma, S. Gupta, H. Shin, et al., "GeoEngine: A platform for production-ready geospatial research," in *CVPRD*, 2022, pp. 21416–21424.