GENERATION OF SIMULATED "ULTRA-HIGH RESOLUTION" HIRISE IMAGERY USING ENHANCED SUPER-RESOLUTION GENERATIVE ADVERSARIAL NETWORK MODELING. E. Clabaut[1], M. Lemelin[1] and M. Germain[1], [1]Département de Géomatique Appliquée, Université de Sherbrooke, Sherbrooke, Canada, J1K 2R1 (etienne.clabaut@usherbrooke.ca).

Introduction: Convolutional neural networks have been used ever since the initial successes obtained in computer vision in a lot of different applications such as object detection, image recognition, semantic segmentation or super-resolution [1–3]. Single image super-resolution (SR) consists in increasing the resolution of a low resolution (LR) image to a higher resolution (HR) image. Generative adversarial networks (GANs) proved to be efficient in SR as demonstrated in [3] with Enhanced-SRGAN (ESRGAN). This architecture is now widely used in satellite imagery or in model prediction enhancement (*e.g.* [4, 5]). Training GANs consist in training a generator network to fool a discriminator network. In the case of SR, a HR image is downscaled to a lower resolution. The generator tries to build a fake HR image that could not be distinguished from the real one by the discriminator. This method implies that the model will only reconstruct HR images that could be already available. What is the point? In this study, we show that an ESRGAN model, can be used to generate simulated HiRISE images of higher spatial resolution than the original with high fidelity. The new details are "hallucinated" by the model. Other studies (*e.g.* [6, 7]) did SR for Mars imagery but to the best of our knowledge, this study is the first to use GANs on single image SR.

Methods: First, we built our dataset using 49 HiRISE color images of 25 to 50 cm of spatial resolution. These images cover a wide range of geomorphological features such as dunes, craters, canyons or ancient riverbeds, which are used to show the model the textures that could be encountered on Mars. We cut each image in tiles of 1024 by 1024 pixels resulting in a total of 3520 tiles. We used a bicubic downscaling operation with a scale factor of 2 to obtain 512 by 512 pixel image tiles of 0.5 to 1 m of spatial resolution. The resulted dataset is called "DownScaled 2" (DS 2). This was done a second time. The resulted dataset, with 256 by 256 pixels images (1 to 2 m of spatial resolution) is called "DownScale 4" (DS 4). 1200, 150 and 150 images tiles were randomly chosen for the training dataset, the validation dataset and the test dataset respectively. It is important to note that the test dataset is not seen by the model during the training phase. An ESRGAN model is trained to reconstruct the DS 2 dataset from the DS 4 dataset. Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) are used to quantify the degradation between the real and the fake tiles on the test datasets only. Without new training, the same model is used to reconstruct HR images from the DS 2 dataset. Again, PSNR and SSIM are used to quantify the quality of the reconstructed images. Finally, a new model is trained to build HR images from the DS 2 dataset and then, used to construct images with higher resolution than the original HiRISE dataset. The overall methodology is presented in figure 1.

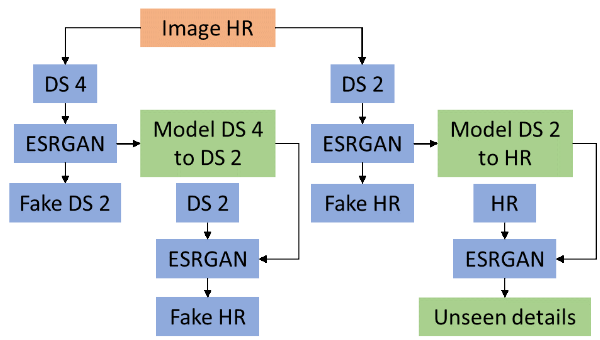

Figure 1. Methodology used to get unseen details on HiRISE imagery.

Results: In this section, 300 PSNR and SSIM values obtained from the two training phases are presented. The model trained to go from the DS 4 dataset to the DS 2 dataset obtained an average PSNR value of 43.76 with a standard deviation of 2.62. The averaged SSIM value is 0.97 with a standard deviation of 0.03. Without any new training, the same model obtains an average PSNR value of 44.58 with a standard deviation of 2.29. The averaged SSIM value is 0.97 with a standard deviation of 0.02. Results are presented in the figure 2.

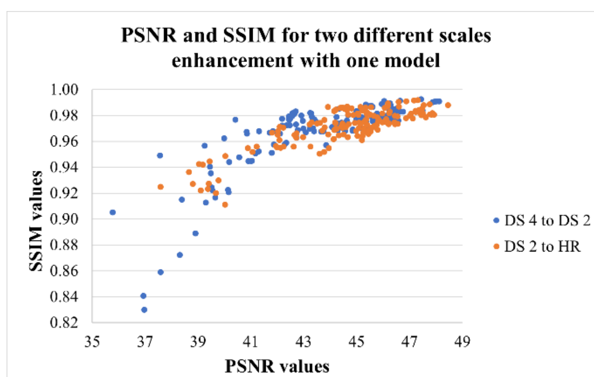

Figure 2. Distribution of PSNR and SSIM values obtained for one model applied on DS 4 dataset and DS 2 dataset with a scale factor of 2.

Discussion: The PSNR and SSIM values presented in the previous section were obtained with only one model, trained to reconstruct images from 1 to 2 m of spatial resolution to 0.5 to 1 m of spatial resolution. This model was applied on the DS 4 dataset and obtained a PSNR value of 43.76. Surprisingly, the same model obtained a slightly higher value, 44.58 on the DS 2 dataset without being trained on it. However, the difference stays inside the standard deviation and should not be considered as significant. Also, there was no decrease in SSIM average value. Hence, no deterioration should be noticed on the tiles. This can be visually confirmed on the images themselves (Figure 3) as they are not visually distinguishable.

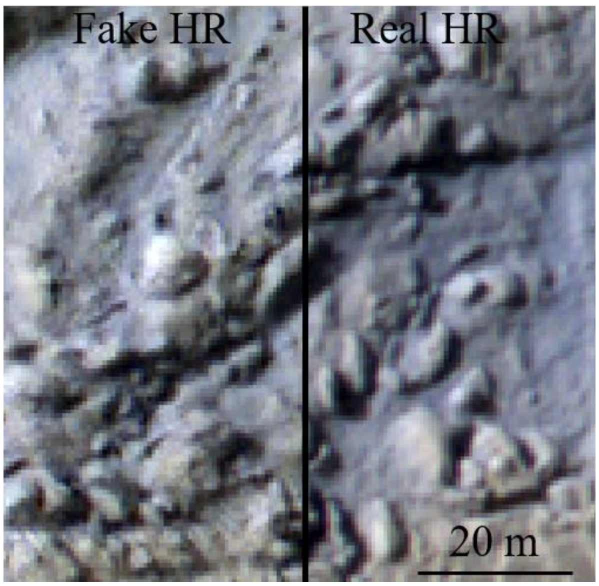

Figure 3. Comparison between a fake HR tile, reconstructed from the DS 2 dataset and the original tile. Tile cut from the "ESP_066239_1115_COLOR" image.

These results suggest that, if a model is trained from the DS 2 dataset to reconstruct HR images, this same model could provide "hallucinated" details with high quality. This statement is valid, assuming that there is no significative change in the ground texture when going from the DS 2 dataset to the HR dataset. Our results support this assumption. The figure 4 shows an ultra-high resolution (UHR) image with "hallucinated" details. The image is highly convincing with much sharper outlines and no artifact that could be noticed. In this case, no PSNR or SSIM can be calculated as no ground truth exists. Hence, the overall quality can be inferred visually only. Still, without any objective quality assessment of the "hallucinated" tiles, these UHR data should be considered as reliable.

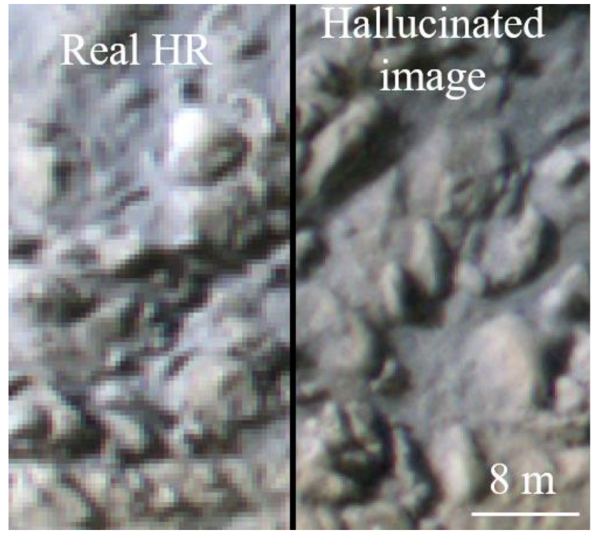

Figure 4. Comparison between the original image and the hallucinated one with unseen details.

Conclusion: Two downscaling steps, each by a factor 2, were applied on HiRISE images. An ESRGAN model trained to reconstruct images from the downscale 2 to the downscale 1 performed very well to reconstruct images from the downscale 1 to the original dataset resolution. This strongly suggests that a model trained to reconstruct images from the downscale 1 to the original dataset can be used to build reliable UHR data. Our test results are good enough to expect that even an enhancement factor of 4 could be used. Such a method would allow all the HiRISE data to be rescaled to an astonishing 12 - 13 cm of spatial resolution whatever the original data spatial resolution between 25 to 50 cm. The method used herein could also be used to enhance the spatial resolution of Context Camera images from its 6 m native spatial resolution to 1.5 m, where HiRISE data is not available.

Acknowledgments: We thank gratefully the NASA and the University of Arizona for providing all the HiRISE data for free.

References: [1] Dong C. et al. (2014) in Fleet D. et al. (Ed.) *Computer Vision – ECCV 2014*, *8692*, 184-199). [2] Kim J. et al. (2016) *2016 IEEE Conf. on Comp. Vision and Pattern Recog.*, 1646-1654. [3] Wang X. et al. (2018) *arXiv:1809.00219 [cs.CV]*. [4] Watson C.D. et al. (2020) *arXiv:2012.01233 [physics.ao-ph]*. [5] Wu Z. and Ma P. (2020) *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B3-2020*, 351–356. [6] Kwan C. et al. (2018) *Remote Sensing*, *10*(9), 1416. [7] Tao Y and Muller J.-P. (2018) *in Bruzzone L. and Bovolo F. (eds.) Proc. of SPIE, 10789*, Image and Signal Processing for Remote Sensing XXIV, 1078903**.**