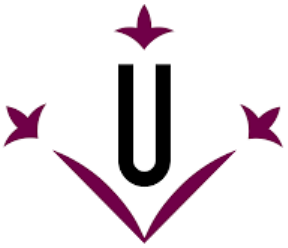

Universitat de Lleida

Economic Viability

Made by
Oriol Alàs Cercós

Delivery
25th of May, 2022

Universitat de Lleida

Escola Politècnica Superior

Màster en Enginyeria Informàtica

Technological Business Management and Entrepreneurship

Professorate:

Josep Escribà Garriga

Contents

List of Figures

Acronyms

CNN Convolutional Neural Network.

GAN Generative Adversarial Network.

NLP Natural Language Processing.

RS Remote Sensing.

SAR Synthetic Aperture Radar.

1 Introduction

Remote Sensing (RS) imagery is critical to perform challenges such climate change or natural resources management, including zone monitoring for reforestation, disaster mitigation and land surface change detection. Nevertheless, on average 55% of the Earth's land surfaces is covered by clouds, being then a significant impediment to carry out a broad range of applications. Satellite imagery plagued by films of clouds that obstructs the scene implies a great loss of information or causing effects such as blurring, which mitigates the power of RS. Hence, RS applications definitely needs a generic technique to detect and remove the cloudy region with an in-painting of the underlying scene.

- Generic end-to-end model that can be retrained for specifically ROI by running supplementary iterations but general enough to be great in all geographical locations.
- Free.

1.1 Sentinel2

Sentinel2 images are provided by two satellites, Sentinel 2A and Sentinel 2B, which orbit each other with a 180^{o} phase shift. The acquisition of the images is 10 days per satellite or 5 days altogether. Therefore, a new updated image of a specific area is available in periods of time not exceeding five days. This makes Sentinel-2 data an excellent choice for studying environmental challenges. Sentinel-2 data is multi-spectral with 13 bands in the visible, near-infrared, and short-wave infrared spectrum. These bands come in a different spatial resolution ranging from 10m to 60m, so the images can be classified as medium-high resolution.

1.2 Related work and state-of-the-art

Deep learning have been a popular and efficient technique to solve challenges from satellite imagery. Specifically, Convolutional Neural Network (CNN) have been the main architecture of neural networks to provide a solution from image-based problems.

In [1], they create a deep learning approach to Sentinel-2 super-resolution. Their hypothesis was the existence of a complex mixture of correlations across many spectral bands

over a large spatial context. Hence, the input of the model is a concatenation of the high-level resolution bands with the low-level resolution bandwidths upsampled to 10m by simple bi-linear interpolations. The model itself is a clear reference of residual networks [2]. Furthermore, as in ResNet architectures, it uses skip connections to reduce the average effective path length through the network, alleviate the vanishing gradient problem and greatly accelerates the learning.

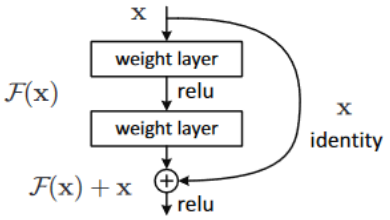

Figure 1: Residual learning: a building block.

Similarly, in [3], a residual network is used with the same skip connections mechanism to bring a solution to cloud removal challenge. It also uses Synthetic Aperture Radar (SAR) optical data, which represents an important complementary source to help the model to make greater results. However, SAR images are affected by a particular type of noise called speckle, which can difficult network's learning. Moreover, the model depends on one more source to remove the clouds, as SAR images cannot be downloaded in Sentinel-2. Regarding the design of the neural network, DSen2-CR is a fully convolutional network, so it can accept input images of any spatial dimensions (m), as it can be seen in 2. The output of DSen2-CR is a 13-channel layer, representing the thirteen bands from Sentinel2.

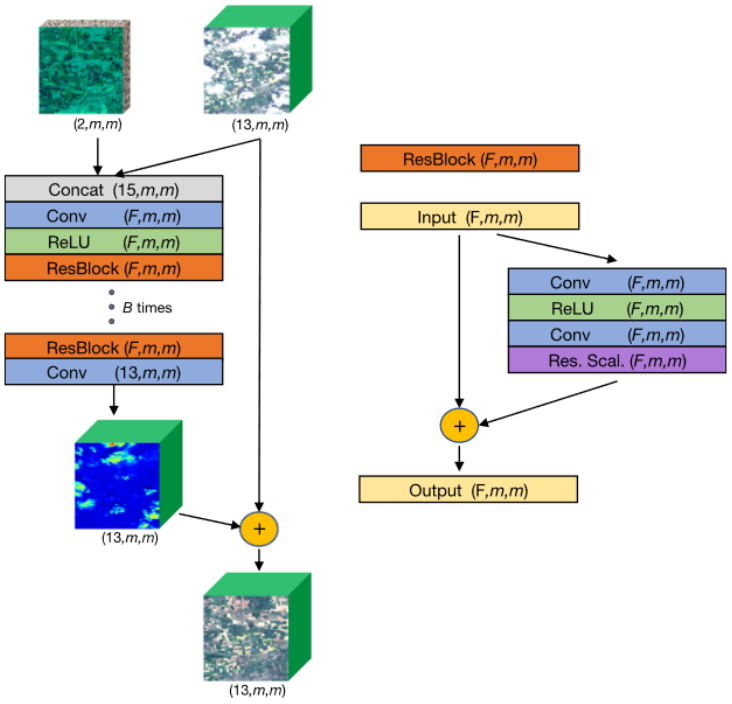

Figure 2: DSen2-CR model diagram.

Improvements have been found using Generative Adversarial Network (GAN) [4], which is a model architecture that belongs to the set of generative models. GAN is an unsupervised model made up of two neural networks. The idea is based on a game theoretic scenario in which the generator network must compete against an adversary. While the generator network produces samples, the aim of the discriminator is to distinguish between the real samples and the drawn by the generator.

During training, both networks constantly try to outsmart each other, in a zero-sum game. At some point of the training, the game may end up in a state that game theorists call a Nash equilibrium, when no player would be better off changing their own strategy, assuming the other players do not change theirs. GANs can only reach one possible Nash equilibrum, which is when the generator produces so realistic images that the discriminator is forced to guess by 50 % probability that the image is real or not. Nevertheless, the training process not always can converge to Nash equilibrium. There are several factors that make the training hard to reach the desired state. For instance, there is a possibility that the discriminator always outsmarts the generator so that it can clearly distinguish between fake and real images. As it never fails, the generator is stuck trying to produce

better images as it cannot learn from the errors of the discriminator. Possible solutions can be carried out such as making the discriminator less powerful, decreasing the learning rate or adding noise to the discriminator target. Another big obstacle is when the generator becomes less diverse, and it learns only to perfectly generate realistic images of a single class, so it forgets about the others. This is called mode collapse. At some point, the discriminator can learn how to beat the generator, but then the latter is forced to do the same but in another class, cycling between classes never becoming good at any of them. A popular technique to avoid is experience replay, which consists in storing synthetic images at each iteration in a replay buffer. There is a lot of literature of obstacles and solutions to improve GAN training and it is still very active, as it is in its applications too.

Regarding cloud removal, in [5], Enomoto proposed a Multi-spectral Conditional Generative Adversarial Network (McGAN) trained using RGB and Near Infra Red (NIR) cloud-free bands as input and synthetic RBG cloudy images as target. It is true that short wave bands are unaffected by cloud cover and using NIR images to guide to uncover the clouds of satellite imagery is great since NIR bands posses have higher penetration through fog than visible light bands. Nevertheless, synthetizing the target might not be realistic enough to feasibly deploy the model in real-state.

Cloud-GAN [6] is a cyclic GAN which uses two generators and two discriminators. Cyclic GANs can create new samples of output data, but also transforming the desired data to samples of input data. In essence, they learn to transform data from the two sources by the two generators respectively. In this case, the generator G_A is generating cloud-free images from cloudy images while the generator G_B is turning the cloud-free images into cloudy. Hence, there is no need to train the model by paired cloudy-cloud-free imagery[1].

Training a CycleGAN using only the two network losses does not guarantee that cycle consistency is held. Thus, an additional cycle consistency loss is used to enforce this property. This loss is defined as the absolute value difference (L1 norm) between an input value x and it's forward-cycle prediction F(G(x)), as well as the input values y and their forward-cycle prediction G(F(y)). The higher the difference, the more distant the

[1]Paired cloudy-cloud-free dataset means to have images from the same zone with and without clouds in order to compare the output data with the cloud-free data.

predictions are from the original inputs. Ideally, our network would minimize this loss.

Although CNN have worked very well for cloud removal, latest and disruptive state-of-the-art deep learning attention-based architectures [7] uncover new paths to achieve remarkable improvements and results. It has been demonstrated that transformers can excellently overcome challenges such Natural Language Processing (NLP) [8] , Text-To-Image Generation [9] or Image Completion [10] with large datasets, great model size and enough compute. Recent contributions have been demonstrated that

2 Data

2.1 Academic Datasets

3 Model architecture

4 Experiments & results

5 Conclusions

6 Bibliography

References

[1] Charis Lanaras, José Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:305–319, 2018.

[2] 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*, pages 770–778, 2016.

[3] Andrea Meraner, Patrick Ebel, Xiao Xiang Zhu, and Michael Schmitt. Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:333 – 346, 2020.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[5] Kenji Enomoto, Ken Sakurada, Weimin Wang, Hiroshi Fukui, Masashi Matsuoka, Ryosuke Nakamura, and Nobuo Kawaguchi. Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1533–1541, 2017.

[6] Praveer Singh and Nikos Komodakis. Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1772–1775, 2018.

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 6000–6010, 2017.

[8] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. 2020.

[9] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8821–8831. PMLR, 18–24 Jul 2021.

[10] Mark Chen, Alec Radford, Rewon Child, Jeffrey Wu, Heewoo Jun, David Luan, and Ilya Sutskever. Generative pretraining from pixels. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1691–1703. PMLR, 13–18 Jul 2020.