Spectral Super–Resolution of Satellite Imagery with Generative Adversarial Networks

Daniel Rojas Pérez

Abstract– Hyperspectral (HS) data is the most accurate interpretation of surface as it provides fine spectral information with hundreds of narrow contiguous bands as compared to multispectral (MS) data whose bands cover bigger wavelength portions of the electromagnetic spectrum. This difference is noticeable in applications such as agriculture, geosciences, astronomy, etc. However, HS sensors lack on earth observing spacecraft due to its high cost. We propose generative adversarial networks as a general-purpose solution to spectral super-resolution of satellite imagery. The proposed network learns mapping from an MS input image to an HS output image, generating nearly 20x more bands than the given input. The results are measured with different approaches: visual interpretation, statistical metrics and application to crop classification. We show that we achieve highly usable applications (to do)

Keywords– generative adversarial networks, hyperspectral imaging, multispectral imaging, spectral resolution, super-resolution

Resum– Les dades hiperspectrals (HS) són la interpretació més precisa de la superfície, ja que proporciona informació espectral fina amb centenars de bandes contigües estretes en comparació amb les dades multiespectrals (MS) les bandes cobreixen parts de longitud d'ona més grans de l'espectre electromagnètic. Aquesta diferència es nota en aplicacions com l'agricultura, les geociències, l'astronomia, etc. No obstant això, els sensors HS manquen a la Terra d'observar naus espacials a causa del seu elevat cost. Proposem xarxes contràries generatives com a solució d'ús general a la super resolució espectral d'imatges de satèl·lit. La xarxa proposada aprèn el mapatge d'una imatge d'entrada MS a una imatge de sortida HS, generant quasi 20x més bandes que l'entrada donada. Els resultats es mesuren amb diferents enfocaments: interpretació visual, mètriques estadístiques i aplicació a la classificació de cultius. Mostrem que aconseguim aplicacions altament utilitzables (per fer)

Paraules clau– generative adversarial networks, hyperspectral imaging, multispectral imaging, spectral resolution, super-resolution

✦

1 Introduction

EARTH observation data is understood as the gathering of information of the surface and atmosphere of earth from a high altitude through remote sensing technical procedures via sensors built in satellites. These sensors often acquire information not only from the three main visible wavelength bands but from finer spectral resolution (width of each band of the spectrum), covering also near infrared (NIR) and short-wave infrared (SWIR) wavelength bands that offer unique remote sensing capabilities. This spectrally different resolution data is classified into multispectral (MS) and hyperspectral (HS) data. HS imaging uses continuous and contiguous ranges of wavelengths (e.i. 400 – 1100 nm in steps of 1 nm) whilst MS imaging uses a subset of targeted wavelengths at chosen locations (e.i. 400 – 1100 nm in steps of 20 nm). There are a handful of MS sensors (i.e., EO–1 ALI, Landsat 7 ETM+, Landsat 8 OLI) covering most of the surface of the earth with high spatial and temporal resolution. Alternatively, HS sen-

• E-mail de contacte: daniel@rojas.ai
• Menció realitzada: Computació
• Treball tutoritzat per: Dr. Felipe Lumbreras Ruiz (Departament de Ciències de la Computació)
• Curs 2020/21

sors (i.e., EO–1 Hyperion) are fewer and only cover specific small areas due to the scenes' small swath, with poor spatial and temporal resolution. In order to support the shortage of hyperspectral imaging, imaging spectroscopy transformation techniques can be proposed for producing imaging that can be used for direct applications such as crop classification and indirect applications such as helping in the design stage of new sensors by assessing or evaluating the spectral and spatial characteristics.

Most of the work involving super–resolution has been in the spatial and temporal modalities in which Deep Learning has achieved state-of-the-art results [?, ?]. There is literature on related lines of research, such as performing data fusion of MS and HS [?] or single hyperspectral image super-resolution [?]. Nevertheless, little study has been involved with increasing the spectral information obtaining a hyperspectral image from a multispectral image. However, a few studies have proposed solutions to this specific task using diverse techniques [?] such as a spectral reconstruction approach [?], spectral resolution enhancement method [?], a pseudo-HS image synthesis algorithm [?] and a latter extended pseudo-HS image transformation algorithm [?]. All of these methods are linearly structured models and do not consider nonlinearity relationships between MS and HS bands. Thus, we will base our study on proposing a non-linear deep learning approach in order to solve this task. Recently, a study proposed Convolutional Neural Network Regression (CCNR) [?] for the transformation of multispectral data to quasi-hyperspectral data which will be implemented in this study as means of a reference for comparison of our proposed approach.

The principle contribution of our study is based on using top-notch deep learning techniques to get the best possible results on spectral super–resolution. We focused on the implementation of Pair-Identical Image-to-Image Translation using Generative Adversarial Networks (pix2pix) and proposing an enhanced architecture especially suited for the huge increasing of number of channels between the input image and the output image which has not been proposed till date. Also, we will extend our work to crop classification of unseen data independent of the crop type that the model has been trained on.

2 Remote Sensing Datasets

2.1 Train and test data for model building

Data acquisition. The remote sensing data for building the model are ALI and Hyperion imagery. Both land imaging instruments are onboard the NASA EO–1, which partially collects data over the same area at the exact same time. ALI, the MS sensor, provides image data from 10 spectral visible NIR (VNIR) and SWIR discrete bands. Hyperion, the HS sensor, collects 242 continuous spectral channels ranging from 0.357 to 2.576 mm with a 10-nm bandwidth. Both sensors have a similar spatial resolution, pixel size, of about 30 meters.

ALI and Hyperion spectral, spatial and temporal domain overlapping makes the two sensors the best choice for building a solid model without the variance of external factors affecting the reliability of the correlation that we dealt to map(*).

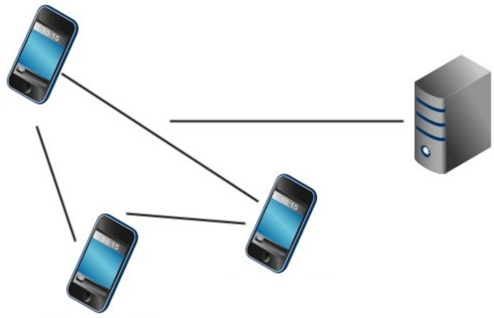

Fig. 1: Multi-Hyper comparison

The product scenes were acquired through USGS Earth Explorer, selecting the L1Gst level of correction, a terrain corrected product provided in 16-bit radiance values. A total of 31 scenes, during a span of 8 years (2008-2016), were collected from the surrounding area of Ciudad Real due to the high-density of land dedicated to crop soil, which produces a higher impact on surface reflectance of SWIR bands[?]. From the 10 ALI bands, 9 of them are MS and 1 is a panchromatic (PAN) band, which we will dismiss for our study due to the fact that it does not add relevant spectral information. From the 242 Hyperion bands, 170 bands (8–57, 79–120, 131–165, 181–223) will be used, the rest are uncalibrated or noisy and would cause a negative impact on the prediction. Hence, 170 will be the number of bands that we will predict from a 9–band input, resulting in a nearly 20x spectral super–resolution.

Data preprocessing. ALI and Hyperion have different product properties, they differ on height, width, pixel size, area covered and such more properties which require some preprocessing before doing any experiment.

Despite being both sensors onboard the same satellite, the pixel to pixel alignment and pixel size are not identical and thus they require to be geometrically corrected. For the correspondence of the scenes, ground control points (GCP) were manually set and ALI scenes were geometrically corrected using first-order polynomial interpolation and bilinear resampling to match Hyperion scenes. Also, their extents were clipped so they have the exact same size and shape and cover the exact same area.

All data is normalized to the [0–1] range and converted to 32-bit float data.

Satellite sensors sometimes deliver a bad behaviour by mis-

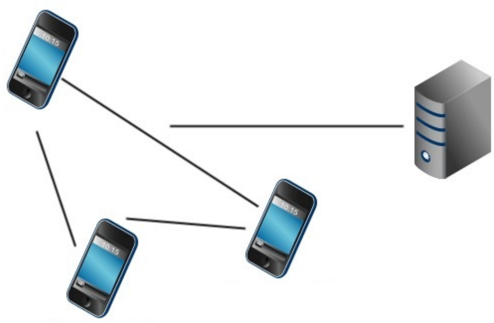

Fig. 2: Patch extraction

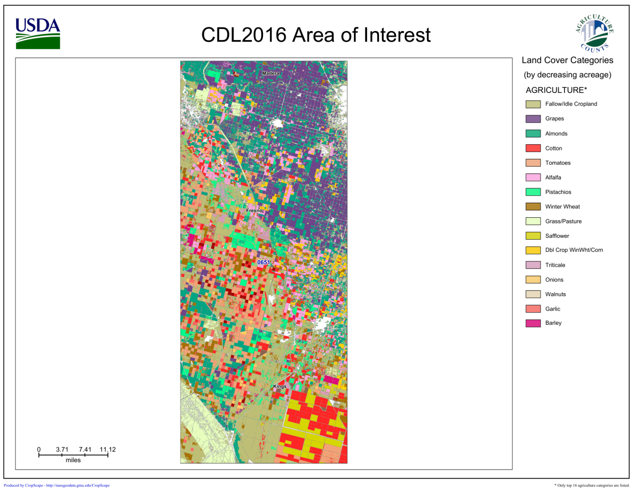

Fig. 3: MS and HS data (change pic)

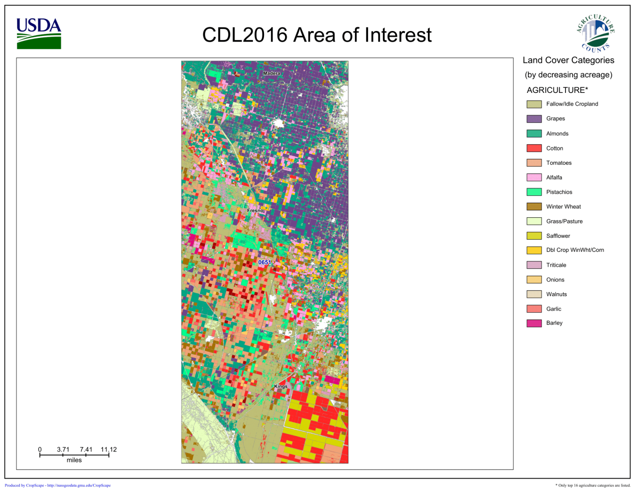

Fig. 4: Crop data in area of interest

sing information in some pixels. During a study of the data, we noticed some spikes on the maximum values of different bands from a single image scene caused by these corrupted pixels. We corrected them by changing their value to the median of the same pixel from the neighbour bands.
The data that will be provided to the model will be in a set of multiple pair-identical patches extracted from the different product scenes, excluding the ones with clouds. The dataset will finally be split in half, 50% of randomly chosen patches for the training set and the rest 50% for the test set.

2.2 Crop classification data

Data acquisition. The image data selected for crop classification is a crop-specific land use data, the Cropland Data Layer (CDL) product from the USDA NASS that covers the entire Contiguous United States (CONUS) at 30-meter spatial resolution with very high accuracy up to 95% for major crop types (i.e., grape, almond) in major crop area.

Data preprocessing. etc

3 Methods

3.1 Generative Adversarial Networks

Generative Adversarial Networks (GANs) are a type of generative model composed of two different networks: the generator and the discriminator. The generator's objective is to generate fake data indistinguishible from real data. On the other hand, the discriminator deals with classifying whether a random sample has been artificially generated or is a real sample data. Both networks will compete during training in which the gain or loss of one of the networks is offset by the gain or loss of the opposite, till the discriminator cannot correctly identify between the samples.

The generator, U-Net, consists of an encoder-decoder architecture in which a convolution is applied after the last layer in the decoder to map the number of output channels, 170 channels in our case.
The discriminator will decide whether the unknown image was generated from the generator or not. The architecture is also called a PatchGAN and will decide if patches from the sub-image patches are real or not.

3.2 Evaluation metrics

The resulting generated imagery will be evaluated through three different evaluation approaches:

Visual interpretation. The visual interpretation and comparison of the results played an important role on the training of the data since we were able to easily interpret whether we had to tune the model because of blurriness issues or offset data. However, it will not be an important decisive factor in the final comparison since the visual interpretation of 170 bands can not be easily readable, although we can have a good glance of it by visualizing three bands composing the RGB channels in a False Color Composite (FCC).

Statistical metrics. Quality of the data will also be statistically evaluated. This evaluation will be separated in two classes: band–wise evaluation and pixel–wise evaluation. Band–wise evaluation will consider the next metrics:

Pearson's Correlation Coefficient (PCC) measures the linear correlation between the real and fake bands

$$FORMULA$$

Root-Mean-Square Error (RMSE) measures the difference between reflectances of real and fake bands

$$FORMULA$$

Peak Signal-to-Noise Ratio (PSNR) measures the quality of the reconstructed image

$$FORMULA$$

Structural Similarity index (SSIM) assesses the visual impact of luminance, contrast, and structure characteristics of the predicted image into a single index metric

$$FORMULA$$

Regarding the pixel-wise evaluation, we used more task-specific metrics which take into account the 170-long vector of reflectances of a single pixel. The next metrics will be measured for the generated data:

Spectral Angle Mapper (SAM) evaluates the difference between two spectra (real and fake pixel) by measuring the angle.

$$FORMULA$$

Spectral Information Divergence (SID) is an information theoretic spectral metric which considers each pixel as a random variable and uses its spectral histogram to define a probability distribution. The spectral similarity between two pixels is measured by the discrepancy of probabilistic behaviours between their spectra.

$$FORMULA$$

Crop classification. etc

4 Experiments and Results

etc

5 Conclusions

etc

Acknowledgment

etc

Referències

[1] http://en.wikibooks.org/wiki/LaTeX

[2] Referència 2

[3] Etc.

Apèndix

A.1 Secció d'Apèndix

... .

A.2 Secció d'Apèndix

... .