
This is the **published version** of the bachelor thesis:

Rojas Pérez, Daniel; Lumbreras Ruiz, Felipe, dir. Spectral Super-Resolution of Satellite Imagery with Generative Adversarial Networks. 2021. (958 Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/238422>

under the terms of the  license

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. In this study, we propose a novel loss function for generative adversarial networks as a spectral-oriented and general-purpose solution to spectral super-resolution of satellite imagery. The proposed architecture learns mapping from MS to HS data, generating nearly $20\times$ more bands than the given input. We show that we outperform the state-of-the-art methods by visual interpretation and statistical metrics.

Keywords– generative adversarial networks, hyperspectral imaging, multispectral imaging, spectral resolution, super-resolution, spectral angle mapper, loss function

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 és notable en àmbits com l'agricultura, les geociències, l'astronomia, etc. No obstant això, els sensors HS manquen a les naus espacials d'observació terrestre a causa del seu elevat cost. En aquest estudi proposem una nova funció de pèrdua per a generative adversarial networks com a solució orientada a l'espectre i de propòsit general per la superresolució espectral d'imatges de satèl·lit. L'arquitectura proposada aprèn el mapatge de dades MS a HS, generant gairebé $20\times$ més bandes que l'entrada donada. Mostrem que superem els mètodes state-of-the-art mitjançant la interpretació visual i les mètriques estadístiques.

Paraules clau– generative adversarial networks, imatges hiperspectrals, imatges multiespectrals, resolució espectral, superresolució, spectral angle mapper, funció de pèrdua



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 (i.e. continuous 1 nm channels in the range 400–1100 nm) whilst MS imaging uses a subset of targeted wavelengths at chosen locations (i.e. specific 20 nm channels in the range 400–1100 nm). There are a handful of MS sensors (i.e., EO-1 ALI [1], Landsat 7 ETM+, Landsat 8 OLI) covering most of the surface of the earth with high spatial and temporal resolution. Alternatively, HS sensors (i.e., EO-1 Hyperion [2]) are fewer and only cover specific small areas due to the scenes' small swath, with poor spatial and temporal resolution.

- 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

2 RELATED WORK

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 [3, 4]. There is literature on related lines of research, such as performing data fusion of MS and HS [5] or single hyperspectral image super-resolution [6]. 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 [7] such as a spectral reconstruction approach [8, 9], spectral resolution enhancement method [10], a pseudo-HS image synthesis algorithm [11, 12] and a latter extended pseudo-HS image transformation algorithm [13]. 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 nonlinear deep learning approach in order to solve this task. Recently, a study proposed Convolutional Neural Network Regression (CCNR) [14] 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 principal 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 multi-image and single-image spectral super-resolution in order to extract a solid analysis.

3 REMOTE SENSING DATASETS

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 μm 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.

The product scenes were acquired through USGS Earth Explorer, selecting the L1Gst level of correction, a terrain

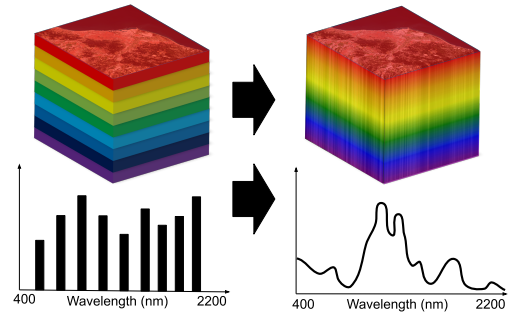


Fig. 1: Comparison of MS and HS data.

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 [15]. 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 $20\times$ 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 missing 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

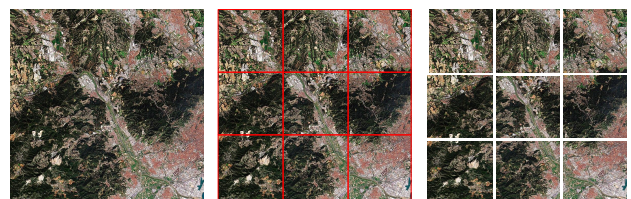


Fig. 2: Patch extraction example.

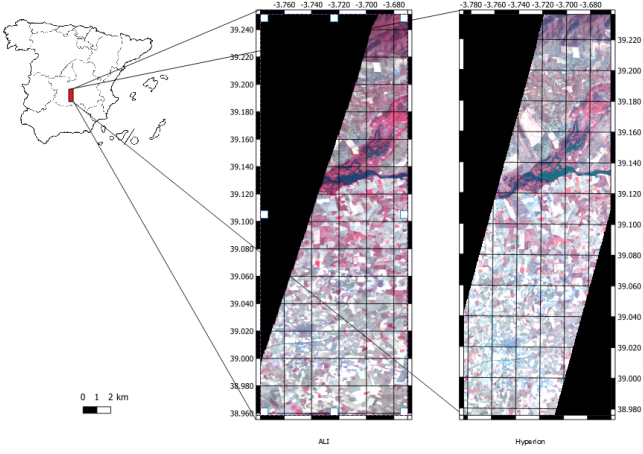


Fig. 3: Overlapping ALI (left) and Hyperion (right) data. ALI RGB: (6,5,4). Hyperion RGB: (45,21,14)

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.

Single-Image model data The area for performing Single-Image Spectral Super-Resolution analysis has been selected through the USDA NASS Cropland Data Layer (CDL), a crop land crop-specific land use data that covers the USA, visualized in Figure 4. We decided to define the area of interest based on the high crop density of the area. Same preprocessing to ALI and Hyperion as explained in last section have been realized.

4 METHODS

4.1 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 indistinguishable 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.

We will focus our study around the Pix2Pix [16] model, a state-of-the-art framework for image translation based on GANs extracting correspondence features between pairs of images. This model is a general-purpose solution for image translation tasks and is sometimes limited due to this generalization, which makes it non-specialized in capturing relationships between original and constructed images of spe-

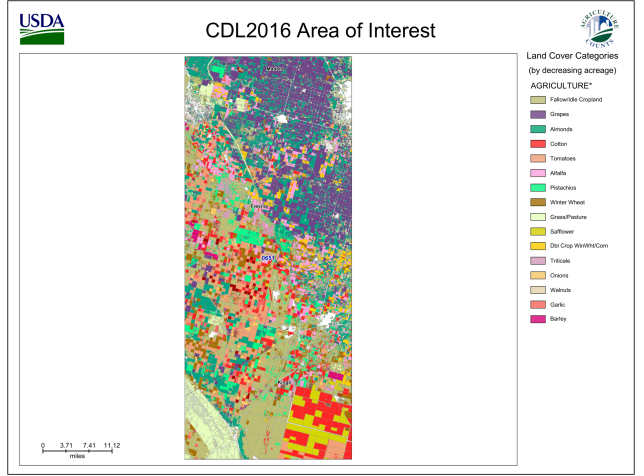


Fig. 4: Crop type data in the area of California, USA.

cific constraints, characteristics and abstractions.

Transforming Multispectral data to Hyperspectral data is a clear example of a task and data that differs from the standard example solutions in which Pix2Pix is often showed performing.

Pix2Pix is based on Conditional Generative Adversarial Network (CGAN), whose generator network learns a mapping from the input image x to the target image y , i.e., $\{x \rightarrow y\}$ and the discriminator network will try to distinguish them if real or fake. The loss function of Pix2Pix is defined by:

$$\begin{aligned} \mathcal{L}_{Pix2Pix}(G, D) = & \mathbb{E}_{x,y}[\log D(x,y)] + \\ & + \mathbb{E}_x[\log(1 - D(x, G(x)))] + \\ & + \lambda \cdot \mathbb{E}_{x,y}[\|y - G(x)\|_1] \end{aligned} \quad (1)$$

4.2 SAM-GAN: Proposed enhancement

Most of the loss functions used in architectures are based on two-dimensional feature extraction, computing the error between data extracting them channel by channel. When dealing with hyperspectral data, the most important features and relationships to aim to be extracted are not only on the spatial dimension but on the spectral dimension too. In this study, we propose a loss function that gathers both high-level features from the spatial dimension and the long spectral dimension in order to achieve a solid learning of the structure of the HS data.

The loss function is based on Pix2Pix’s loss function but adding the Spectral Angle Mapper (SAM) metric to minimize the errors of spectral features. SAM qualifies the similarity of the original and the transformed vector reflectance across the spectra through measuring the average angle between them. The proposed SAM-GAN architecture’s loss function is defined by (2).

$$\begin{aligned} \mathcal{L}_{SAM-GAN}(G, D) = & \mathcal{L}_{Pix2Pix}(G, D) + \\ & + \lambda \cdot \frac{1}{n} \arccos \left(\sum_{i=1}^n \frac{(\mathbf{I}_{gt}^{(i)})^T \cdot \mathbf{I}_{re}^{(i)}}{\|\mathbf{I}_{gt}^{(i)}\|_2 \cdot \|\mathbf{I}_{re}^{(i)}\|_2} \right) \end{aligned} \quad (2)$$

\mathbf{I}_{gt}^i and \mathbf{I}_{re}^i are the i^{th} pixel of the ground truth image and the transformed HS image, respectively; T equals

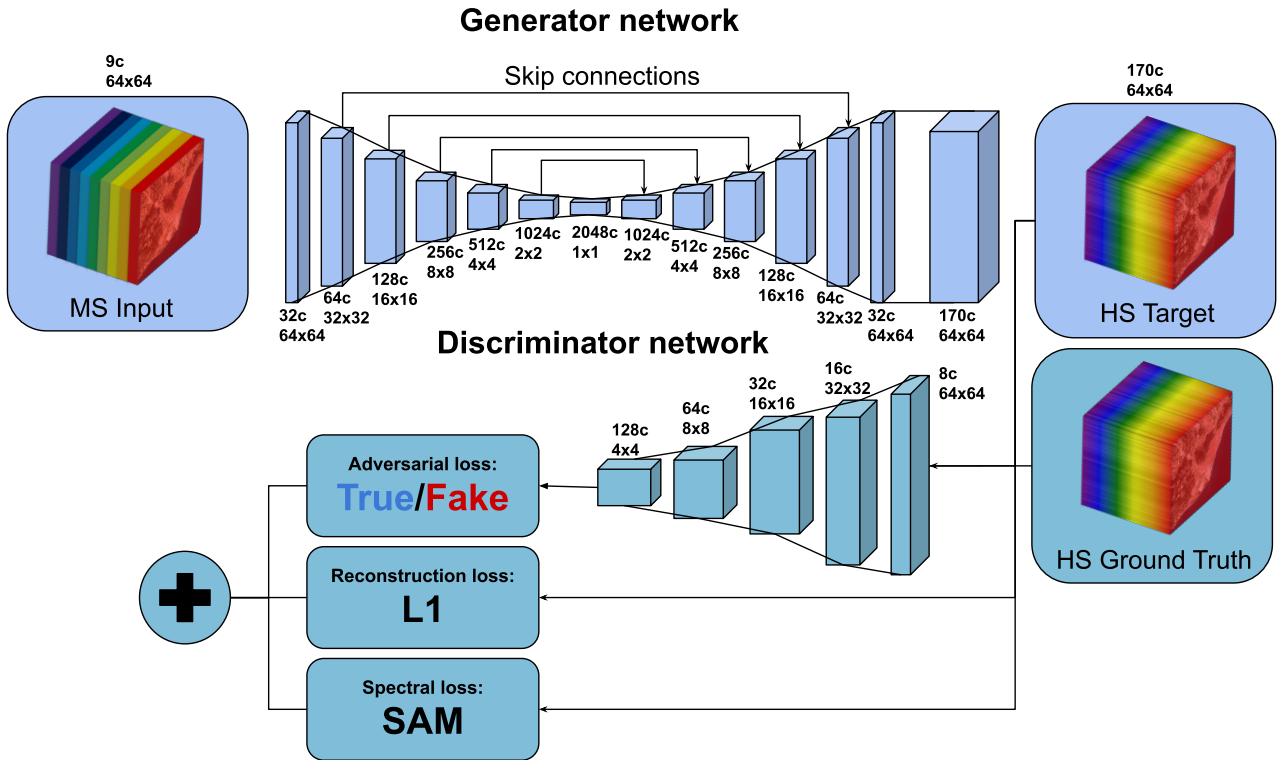


Fig. 5: SAM-GAN architecture

transpose, and n is the total number of pixels of each image.

The architecture is composed by U-Net as the generator and Patch-GAN as the discriminator, the generator will process a batch of 128 image patches and apply a series of 6 contracting blocks, each of them composed by two 2-D convolutional filters with a kernel size of (3×3) are computed on the patches followed by Batch Normalization and a max pool operation (in the first three blocks we also apply dropout of 50%). In each contracting block, the width and height of the patches get split (i.e., $64 \times 64 \rightarrow 32 \times 32$) and the number of channels doubled, achieving on the deepest state of the generator a number of 2048 channels in 1×1 patches. The batch will then get processed by the expanding blocks which performs an upsampling, a convolution, a concatenation with the output of the mirrored contracting block and finally two convolutional layers. 6 contracting blocks in total will set the patches on their original structure and the feature mapping convolutional last layer will match the output desired by our specific task. The discriminator, similar to the contracting path of the generator architecture, will decrease the spatial size of the image while increasing the spectral dimension and will output a matrix classifying whether a minipatch of the image patch generated is real or fake.

For our experiments, we set a learning rate of 0.005 for the first 70 epochs and then decrease it to 0.001 to allow the model to learn a more optimal set of weights. This architecture uses LeakyReLU for the contracting blocks, ReLU for the expanding ones and Sigmoid for the final output of the U-Net. The objective function is, as previously stated, a weighted sum of the discriminator loss, the L1 reconstruction loss multiplied by $\lambda = 200$ and the proposed Spectral Angle Mapper loss multiplied by $\gamma = 200$. Due

to the fact that through the training and tweaking process, our model generated best results when the SAM loss had a higher level of impact on the loss function.

Layer (type)	Output Shape
Conv2d	[-1, 32, 64, 64]
Contracting Block	[-1, 64, 32, 32]
Contracting Block	[-1, 128, 16, 16]
Contracting Block	[-1, 256, 8, 8]
Contracting Block	[-1, 512, 4, 4]
Contracting Block	[-1, 1024, 2, 2]
Contracting Block	[-1, 2048, 1, 1]
Expanding Block	[-1, 1024, 2, 2]
Expanding Block	[-1, 512, 4, 4]
Expanding Block	[-1, 256, 8, 8]
Expanding Block	[-1, 128, 16, 16]
Expanding Block	[-1, 64, 32, 32]
Expanding Block	[-1, 32, 64, 64]
Conv2d	[-1, 170, 64, 64]
Sigmoid	[-1, 170, 64, 64]
Total params:	117,456,074
Trainable params:	117,456,074
Non-trainable params:	0

4.3 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,

$$PCC_i = \frac{\sum_{j=1}^n (p_{ij}^H(\lambda) - \mu_i^H)(p_{ij}^{\hat{H}}(\lambda) - \mu_i^{\hat{H}})}{\sqrt{\sum_{j=1}^n (p_{ij}^H(\lambda) - \mu_i^H)^2} \sqrt{\sum_{j=1}^n (p_{ij}^{\hat{H}}(\lambda) - \mu_i^{\hat{H}})^2}}.$$

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

$$RMSE_i = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}.$$

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

$$PSNR_i = 10 \cdot \log_{10} \frac{R^2}{MSE}.$$

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

$$SSIM_i = \frac{(2\mu_H\mu_{\hat{H}} + C_1)(2\sigma_{H\hat{H}} + C_2)}{(\mu_H^2 + \mu_{\hat{H}}^2 + C_1)(\sigma_H^2 + \sigma_{\hat{H}}^2 + C_2)}.$$

where every metric will be measured for each band having wavelength λ , H and \hat{H} are the real and fake data, $p_{ij}^{\hat{H}}$ represents the reflectance value of real data at pixel j in band i . The mean of all pixels in a single band is represented by μ and n is the quantity of pixels in a band. R equals to the maximum possible value of the image data.

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.

$$\alpha_j = \arccos \frac{\sum_{i=1}^N p_{ij}^{\hat{H}}(\lambda) p_{ij}^H(\lambda)}{\sqrt{\sum_{i=1}^N p_{ij}^{\hat{H}}(\lambda)^2} \sqrt{\sum_{i=1}^N p_{ij}^H(\lambda)^2}}$$

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.

$$SID_j = D(p_{ij}^H | p_{ij}^{\hat{H}}) + D(p_{ij}^{\hat{H}} | p_{ij}^H) \quad (3)$$

where $D(p_{ij}^H | p_{ij}^{\hat{H}}) = \sum_{i=1}^N p_i \log(p_i/q_i)$

5 EXPERIMENTS AND RESULTS

5.1 Model performance comparison

For all the experiments we run each model for 400 epochs and evaluated the initial progress and the final model performance over the test set at the final epoch. Same randomly chosen data is used for the experiments as it has been extracted to be used in a free-environmental external factor, grabbing large portions of data from the same area over the years and randomizing it so it is able to generalize for its application to multiple use cases. We are going to be comparing our base model with the last proposed deep learning approach of the spectral super-resolution task by [14] to appreciate the results of implementing a larger and a more powerful deep learning structure able to learn more advanced features. Furthermore, we will compare with the Pix2Pix model in order to measure how our network enhancement performs directly competing with its basic architecture. In order for it to be a fair comparison, Pix2Pix and SAM-GAN has the exact same architecture structure and the exact same hyperparameter values for the exception of adding to the loss function the weighted Spectral Angle Mapper as defined previously.

Earlier learning than previous methods. The transformed hyperspectral data are generated from the test dataset samples and have been evaluated with all six statistical metrics. The results show a plot with x ranging from 0 to 20 epochs and the y axis shows the metric punctuation at any given epoch.

The statistical measures show that SAM-GAN provides the most superior performance in all spectral-based metrics such as SAM and SID and delivers a fast convergence by getting to better results in a shorter period of time. Regarding band-wise metrics (PCC, RMSE, PSNR and SSIM), our model show best performance at PCC, PSNR and SSIM, getting best results than the other two architectures. RMSE is a band-wise metric specialized in computing 2-dimensional spatial error, this feature is used by CNNR (using MSE as an objective function, hence minimizing it and obtaining better results) and Pix2Pix (with its reconstruction loss) and both models are able to generate faster and better results than our model in the first 10 epochs. Although our model achieves similar results after the first 10 epochs.

The main difference between Pix2Pix and SAM-GAN is the difference in learning time and we get slightly better end-results while requiring the same computational time per epoch. The difference in learning time is due to that Pix2Pix is mainly benefited from the discriminator in order to learn valuable features. This, in counterpart, makes the learning process less stable and longer, since the discriminator plays a huge role and has to learn classifying features too. On the other hand, the SAM-GAN architecture benefits from

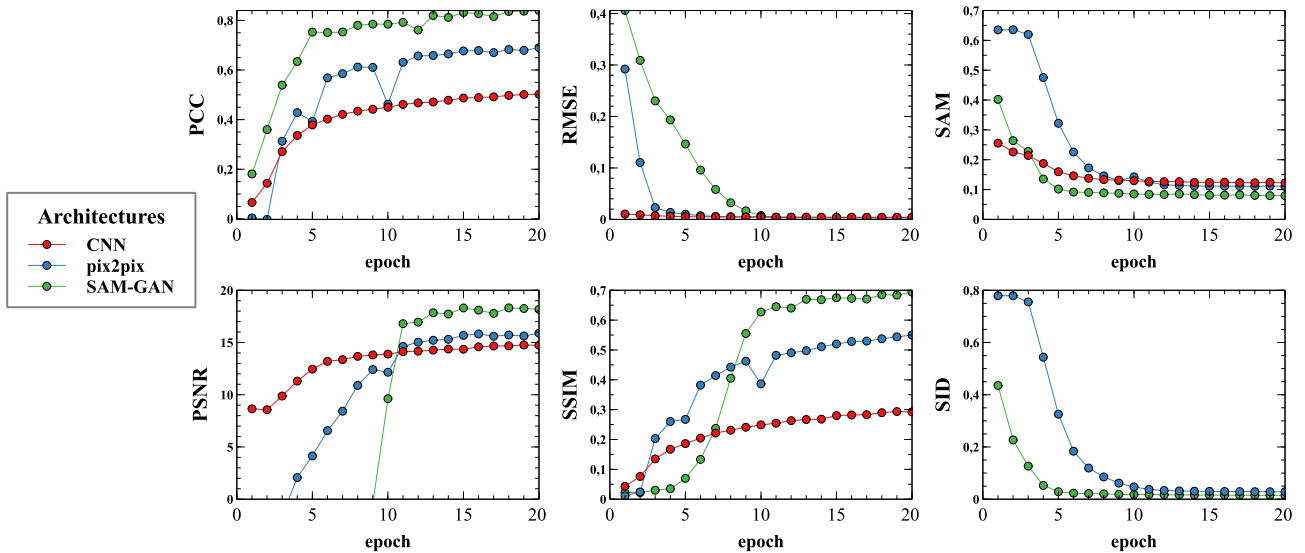


Fig. 6: First 20 epochs of every training process.

PCC, PSNR, SSIM: higher is better. RMSE, SAM, SID: lower is better.

the Spectral Angle Mapper loss by providing instant relationship information between the whole spectral dimension, causing in a rapid convergence.

Visual comparison. Visual comparison shows an even clearer picture of the performance of the three tested models. Same bands have been plotted, conforming a False Color Composite combination in order to show the reconstructed bands. We are only showing 4 generated patches examples of three bands out of 170 generated bands, so the representation is limited but they provide a solid justification of the statistical results. Epochs 4, 9 and 12 are showed in order to appreciate the early learning difference between models. The 400th epoch is shown as it shows the final results of the fully-trained architectures performing at their best. Last row are the Ground Truth data showed for visual comparison as the target result that the architectures were trying to generate.

The visual comparison show great differences between the methodologies of learning between the three different architectures. CNNR starts producing softer, granulated patches due to the loss function that it is trying to minimize and the fewer parameters that it has compared to other architectures, which may cause a longer need of training time to learn the appropriate mapping between multispectral data and hyperspectral data, deducted from the non-variation of the first epochs of training. Pix2pix shows some greater level of detail in the first epochs although there are some parts where the generated sample produced a non-desired prediction. This issue is very common on the first steps of training of Pix2Pix models as the discriminator is also learning to classify between real or fake. After the 9th epoch, the results shown in the second and third row are better than the CNNR competitor, showing fine-grained level of definition and an overall superior structure. The SAM-GAN architecture, our model, shows a very different generated samples from the 4th stage of training but is at the same time a detailed patch with characteristics well-defining the ground truth data. As previously mentioned, SAM-GAN benefits from a very early adaptation to the task and it shows at 9th epoch,

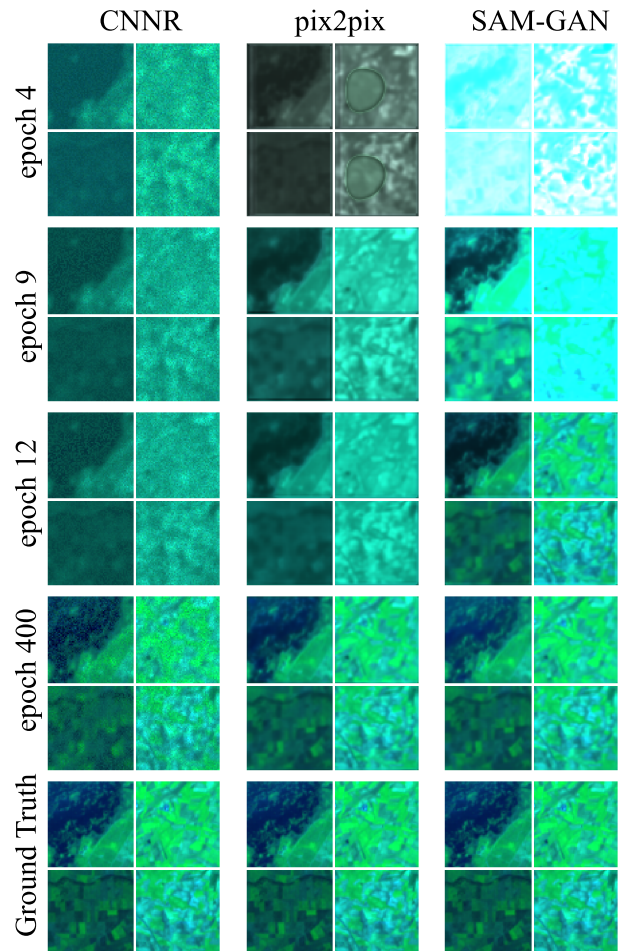


Fig. 7: Generated samples from each architecture at different training steps.

visualizing two of the presented samples with a very high fidelity. After three epochs, at the epoch 12th, SAM-GAN already achieves nearly-perfect results while its competitors still lack level of definition, showing quite different reflectance values comparing to the ground truth patches.

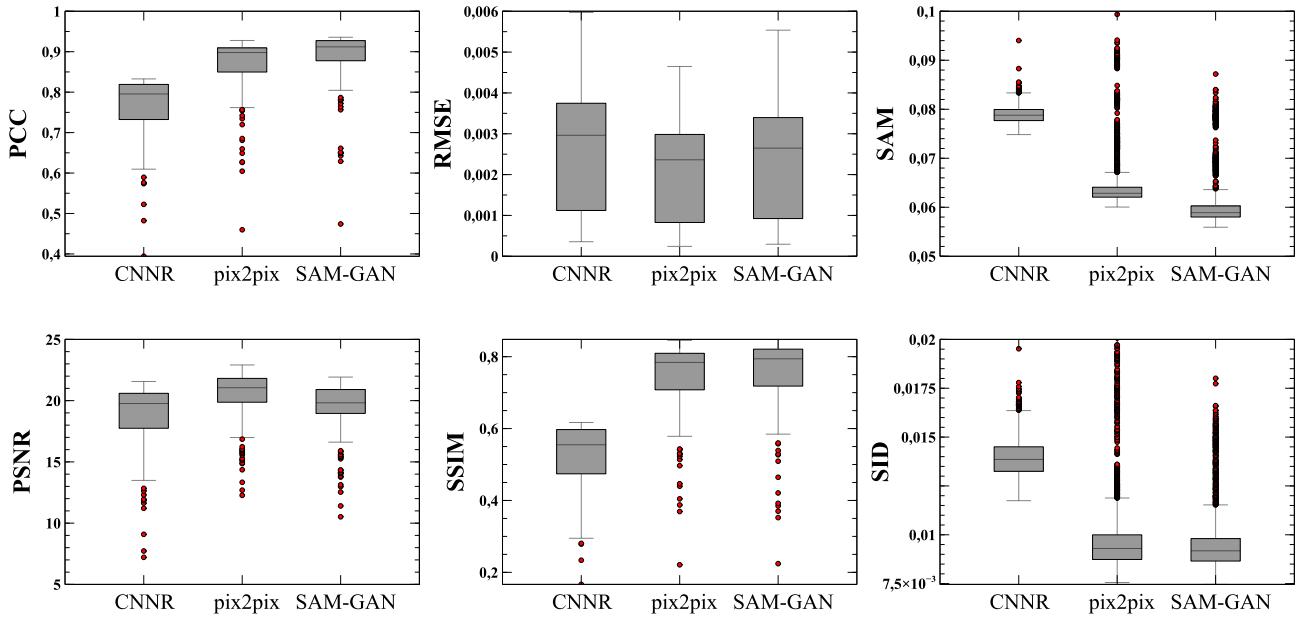


Fig. 8: Models evaluation

Fully-trained model comparison. The generated data is evaluated with six statistical metrics. For the band-wise metrics (PCC, RMSE, PSNR and SSIM), the evaluation is computed for each narrow spectral band, comparing pairs of bands from real and generated data from the test dataset and computing the error. Hence, the boxplot presents the performance of the bands, showing outlier bands that did not get to perform as well as the others. Regarding the pixel-wise metrics, these are computed comparing pixel-to-pixel from real and generated patches and comparing the whole vector of reflectance values of size 170.

It is observed that both Pix2Pix and SAM-GAN outperforms the CNNR architecture in all statistical metrics, due to using more advanced deep learning techniques and being able to extract higher-level features.

Between the two GAN architectures, SAM-GAN is clearly performing better when compared to Pix2Pix on PCC, SSIM and SAM, being the latter the objective function and the evaluation metric it is scoring the best compared to the other models. Both GANs are getting similar performance on RMSE and SID metrics. Due to the fact that Pix2Pix’s L1 loss weights more than SAM-GAN’s L1 loss for the final objective function, SAM-GAN does not outperform on RMSE but is getting a solid score, similar to the other models, and it means that SAM-GAN does not lose the spatial features learning by adding the spectral dimension learning. Our architecture does perform worse on PSNR despite acquiring a very fast learning in that specific metric as shown in Figure 6.

Band-wise metrics per band. It is interesting to observe which are the bands causing a decrease on the evaluation of the generated data independently of the model. Figure 9 shows the specific band (numerated by index in the channel order mentioned in Section 2) performance, being the x axis, the index of the band, and the y axis, the metric score. This evaluation could not apply to SAM and SID due to its nature of pixel-to-pixel comparison, not band-to-band.

Table 1: BAND-WISE METRICS

Architecture	PCC	RMSE	PSNR	SSIM
CNNR	0.767	0.00274	18.69	0.528
Pix2Pix	0.867	0.00215	20.37	0.741
SAM-GAN	0.886	0.00245	19.34	0.753

Table 2: PIXEL-WISE METRICS

Architecture	SAM	SID
CNNR	0.0789	0.01391
Pix2Pix	0.0649	0.00991
SAM-GAN	0.0601	0.00944

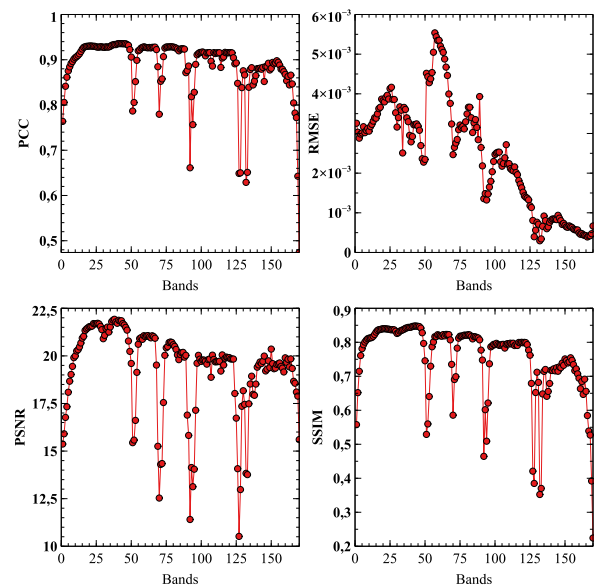


Fig. 9: Band-wise metrics per band.

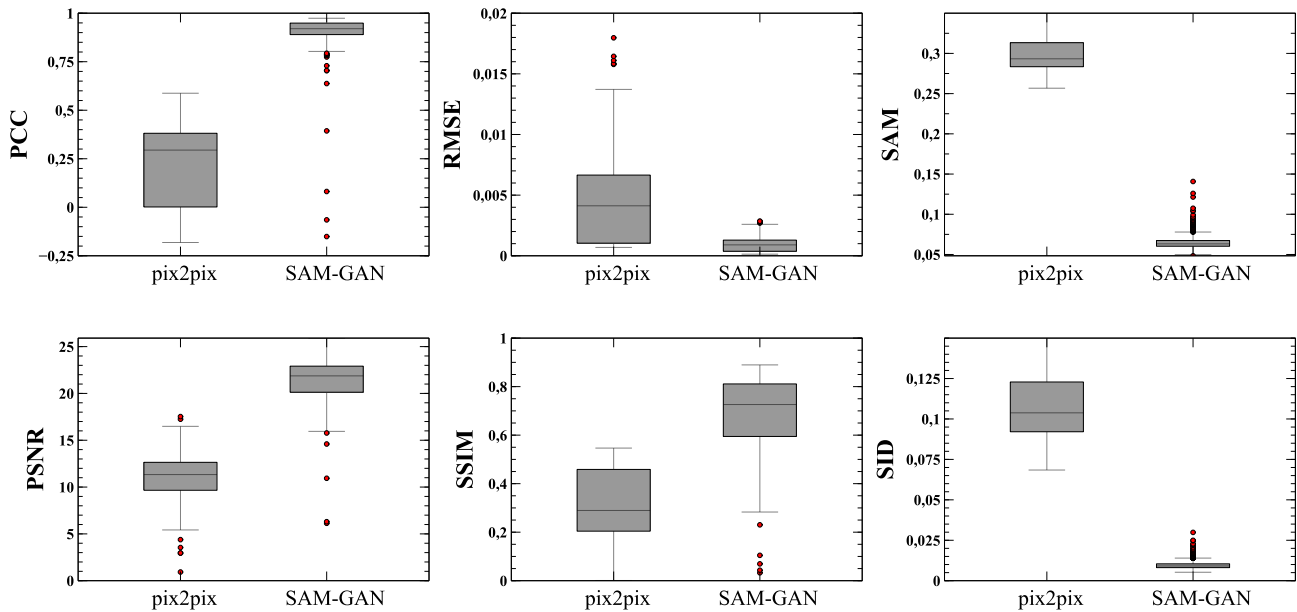


Fig. 10: Models trained and tested in a single scene comparison.

Figure 9 shows that the bands not performing as well in every metric are always, in different measure, the same. The explanation for this phenomenon may be that it is due to a poor band co-registration for specific portions of the wavelength spectrum due to errors or miscalculations of the spacecraft’s sensor.

We should note that, despite those bands having worse performance than the average, it still is achieving very decent results. It depends on the application use whether the worse bands add value to the generated data or not. In some cases, just a handful of principal components are required to perform a classification task or a segmentation task so it wouldn’t be required to generate those bands and is not a big deal to lose that information. In other cases, such as sensor engineering, a more reliable generated data could be required to be helpful at all.

5.2 Single-Image Spectral Super-Resolution

It is interesting to test models when having less amount of data due to a lack of data or computational power to handle it. In the past section, we trained and tested the model with 31 scenes, composing a total of nearly 2000 64×64 patches for the whole dataset. In this section, we will train and test the architecture making use of just 1 scene in the area of California, composing a total of about 80 patches.

For this experiment, batch size of 128 was reduced to 16. The rest of the hyperparameters and model architecture remained untouched and were not tweaked at any moment as means of testing if the models were able to achieve good results without the need of hyperparameter tuning.

It is observable in Figure 10 that SAM-GAN absolutely outperforms Pix2Pix on every metric we evaluated both models on. While SAM-GAN still gets a great performance similar to the model trained on multiple scenes, this time, Pix2Pix fails to map a decent prediction of HS data.

SAM-GAN’s fast learning ability that we remarked on the past section is also noticeable in this experiment where the architecture performs exceptionally.

The characteristics of model building in just one single sce-

Table 3: BAND-WISE METRICS

Architecture	PCC	RMSE	PSNR	SSIM
Pix2Pix	0.219	0.00486	10.79	0.317
SAM-GAN	0.891	0.00101	21.40	0.678

Table 4: PIXEL-WISE METRICS

Architecture	SAM	SID
Pix2Pix	0.3151	0.12253
SAM-GAN	0.0643	0.00950

ne are that the testing patches areas are completely unseen as we do not have patches over the same area from other scenes. The explanation why SAM-GAN is still able to generate great HS data is that our architecture learns the spectral composition of the land. This results in an architecture that does not need to overfit to the spatial data to generate reliable HS data, but will perform greatly for seen spectral compositions.

In fact, slightly better SAM-GAN results were extracted from this experiment compared to the other one. This is because the spectral response of land slightly differs from scene to scene due to environmental factors and physical phenomena such as the angle degree of sun impacting the land, the season of the year, temperature, climate, etc. Since the training and the testing have been realized over the same scene, the spectral response features learned are the same in the model building and will generate better results because of scene adaptation. On the other hand, in the experiment with large data, we were building a generalized model considering all variations of environmental factors.

6 CONCLUSIONS

In this study, SAM-GAN, an enhanced Pix2Pix model, is proposed in order to outperform state-of-the-art methods in the spectral super-resolution task of satellite imagery. The novel specialized loss function benefits from the Spectral Angle Mapper metric to minimize the reflectance spectral vector error pixel-to-pixel when comparing an artificially generated HS image with a real HS image.

Overall, it seems clear that based on the early learning plots, patch visualization and fully-trained statistical metrics evaluation, SAM-GAN achieves a desirable specific enhancement over GAN-based Pix2Pix and other deep learning architectures.

Furthermore, we analyzed how different our network is by changing such a minor aspect of the architecture. By using our own loss function we achieved better results than state-of-the-art methods and completely understands the relationship of reflectances between bands even when trained on a very small dataset and tested on unseen land.

Hence, the objectives established at the initial phase of the project were accomplished with a great degree of satisfaction.

In the near future, this study aims to be expanded exploring the super-resolution of spatial, spectral and temporal dimensions in one single model. This is a task that would not be aimed to replicate the performance of architectures as the one that we proposed but create a general decent solution.

Furthermore, proposing a directly task-related loss function will be a good path on the extraction of both high-level and low-level spectral features by understanding with a more in-depth vision how sensors work and the interpretation of continuous data.

ACKNOWLEDGMENT

The author wishes to thank Dr. Felipe Lumbreras for the guidance and patience through the whole duration of this study and the Computer Vision Center where I remained as an intern. In addition, this work is partially supported by the Spanish Ministry of Science and Innovation under project BOSS TIN2017-89723-P.

REFERÈNCIES

- [1] USGS EROS Archive (2000). url: <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-earth-observing-one-eo-1-ali>.
- [2] USGS EROS Archive (2000). url: <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-earth-observing-one-eo-1-hyperion>.
- [3] Zhang L, Nie J., Wei W., Li Y., Zhang Y. (2020). “Deep Blind Hyperspectral Image Super-Resolution”. In: IEEE TNNLS.
- [4] Jiang J. Sun H., Liu X. et al. (2020). “Learning Spatial-Spectral Prior for Super-Resolution of Hyperspectral Imagery”. In: IEEE TCL.
- [5] Yokoya N. Grohnfeldt C., Chanussot J. (2020). “Hyperspectral and Multi-spectral Data Fusion: A Comparative Review”. In: IEEE GRSM.
- [6] J. Jiang (2020). Hyperspectral-Image-Super-Resolution-Benchmark. url: <https://github.com/junjun-jiang/Hyperspectral-Image-Super-Resolution-Benchmark>
- [7] Yi C. Zhao Y., Cheung-Wai J. (2019). “Spectral Super-Resolution for Multi-spectral Image Based on Spectral Improvement Strategy and Spatial Preservation Strategy”. In: IEEE TGRS.
- [8] V. Tiwari, V. Kumar, K. Pandey, R. Ranade, and S. Agarwal, “Simulation of the hyperspectral data from multispectral data using Python programming language,” Proc. SPIE, vol. 9880, Apr. 2016, Art. no. 98800W.
- [9] V. Tiwari, V. Kumar, K. Pandey, R. Ranade, and S. Agrawal, “Simulation of the hyperspectral data using multispectral data,” in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Jul. 2016, pp. 6157–6160.
- [10] X. Sun, L. Zhang, H. Yang, T. Wu, Y. Cen, and Y. Guo, “Enhancement of spectral resolution for remotely sensed multispectral image,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 5, pp. 2198–2211, May 2015.
- [11] N. T. Hoang and K. Koike, “Development of Bayesian-based transformation method of Landsat imagery into pseudo-hyperspectral imagery,” Proc. SPIE, vol. 9643, Oct. 2015, Art. no. 96430J.
- [12] N. T. Hoang and K. Koike, “Hyperspectral transformation from EO-1 ALI imagery using pseudo-hyperspectral image synthesis algorithm,” Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Tech. Rep. XLI-B7, 2016, pp. 661–665. [Online]. Available: <https://doi.org/10.5194/isprs-archives-XLI-B7-661-2016>
- [13] N. T. Hoang and K. Koike, “Transformation of Landsat imagery into pseudo-hyperspectral imagery by a multiple regression-based model with application to metal deposit-related minerals mapping,” ISPRS J. Photogramm. Remote Sens., vol. 133, pp. 157–173, Nov. 2017.
- [14] Subir P., Nagesh K. (2020). “Transformation of Multispectral Data to QuasiHyperspectral Data Using Convolutional Neural Network Regression”. In: IEEE TGRS.
- [15] Swathandran S, Aslam MAM. “Assessing the role of SWIR band in detecting agricultural crop stress: a case study of Raichur district, Karnataka, India”. Environ Monit Assess. 2019 Jun 16;191(7):442. doi: 10.1007/s10661-019-7566-1. PMID: 31203445.
- [16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. “Image-to-Image Translation with Conditional Adversarial Networks”. CoRR abs/1611.07004 (2016)

APPENDIX

I wished to let this document with the appearance of an academic article and decided to insert the missing content such as project learning objectives, methodology and planification in the appendix.

A.1 Project objectives

This study proposes transforming multispectral data to hyperspectral data in order to increase the spectral information from the MS data by nearly 20 times—from 9 bands to 170 hyperspectral bands—.

The main objectives of this project, along the desired outcome, are presented in the next enumeration:

1. The learning outcomes range from mastering deep learning techniques applied to computer vision, manipulation and processing of satellite remote sensing imagery, learning the state-of-the-art of super-resolution and deep learning in different modalities and be able to use the knowledge to propose novel architectures, performing a large quantity of computational experiments and extract conclusions from analysis.
2. Extreme spectral super-resolution: increasing the number of spectral bands in multispectral data by a factor of over 10.
3. Replicate existing models to compare with and analyzing and understanding their strengths and flaws.
4. Generate architectures that are innovative and competitive in a research area that does not have a large quantity of literature or resources.
5. Achieve domain adaptation and transfer the learning to other scopes of super-resolution such as single-image super-resolution.
6. Propose a novel architecture that achieves state-of-the-art performance

A.2 Methodology

The management of a project is of a great degree of importance to meet the final objectives and deliver them on time. In order to choose the best alternative available, these next points have been taken into consideration:

- Individual project: the work of the project is realized by a single person. Therefore, there is no need of daily communication and coordination with colleagues, neither of a strict schedule.
- Weekly meetings with supervisor: weekly meetings will be held with my supervisor to show the progress and consult any doubts. This gives me the aim to set weekly minor goals and work from the start of the project.
- Poor literature in the area: the research topic has not been deeply studied by other researchers using the techniques that are being planned to be utilized. Thus, there is a level of uncertainty on the overall project,

which makes the timelines not as clear. However, there is some literature on related lines of research, such as performing data fusion of MS and HS [5] or single hyperspectral image super-resolution [6] (which will be analyzed and studied to get a solid understanding of the handling of similar data as ours).

- Novelty: although computer vision and deep learning have already been introduced to me, remote sensing and the satellite scene is a novelty, as well as other tools that will be used. Therefore, the mastering of these tools will be required in order to achieve the goals and it could affect the timelines.

The time management methodology chosen is Kanban¹ due to the fact that it is a methodology characterized for improving the speed and quality of work. It will cause a desire for finishing the current work-in-progress set, as multi-tasking is limited. Moreover, it enables a visible display of the project and work in progress which will be of a great use when communicating with the supervisor. Finally, weekly minor goals will be set in order to increase the motivation and performance.

A.3 Planification

1. Documentation (Weeks: 1st-2nd): study the literature, actual and old, of superresolution and deep learning. We will emphasize on the work related to our field and desired satellites, but we will also explore other alternatives. Moreover, the study of the different characteristics of satellites and their instruments will also be realized. This phase is emphasized in the initial part of the project but will remain active throughout the whole project due to its importance and need.
2. Introduction of new tools (Weeks: 2nd-5th): learning to use new tools such as GDAL [?], QGIS [?], Earth Engine [?], among others, and the downloading and processing of satellite imagery. This phase will be the base to get my further work done.
3. Reproduction of state-of-the-art (Weeks: 3rd-7th): in order to achieve a good understanding of the matter and increase the familiarity with the new tools and methods, a reproduction of the state-of-the-art scene [14] (no code available) will be produced. This phase overlaps with the previous phase due to the fact that this phase is a way of improving the skills needed and achieve a solid learning on the matter.
4. Proposing a novel architecture (Weeks: 7th-13th): more study and computational experiments will be done in order to achieve a novel architecture that stands out from others. This phase is the most exploratory one, it will be based on research among different areas, learning what can be done in order to improve a solution and thinking outside the box. Getting a very broad knowledge, design architectures based on the characteristics that will get the most out of it and performing trial-and-error will be the objectives of this phase. The

¹Trello is the desktop web application chosen for managing the project with Kanban methodology.

timeline for this phase starts from when a good practical understanding is achieved on the problem and it ends soon enough to be able to propose new approaches, such as the zero-shot learning.

5. Single-Image learning approach (Weeks: 13th-17th): the last main goal of the project is that the network can be applied to using very little sets of data, we hope to achieve satisfactory results. This phase remains as the last technical phase, the project could diverge to some different specific goal if we find a viable solution to an unsolved problem.
6. Elaborating a final report (Weeks: 18th-22nd): a final article report will be written up explaining our approach and solution to the problem, specifying the processing of data, the design of the architecture, decisions taken, etc.