POLITECNICO DI TORINO

Master's Degree in Mechatronic engineering

Image processing algorithms for synthetic image resolution improvement

Candidate: Francesco Salvetti
Supervisor: Prof. Marcello Chiaberge

1. General framework and objective

The incessant world population growth of the last century is creating the need for a new way to think to agriculture. Food request is constantly increasing and with it the need of lowering costs and increase production efficiency. One of the most promising idea is to use already available satellite images to have a constant source of information for automatic monitoring of fields parameters. The development of this technology would save the costs for frequent drone flies to collect field shots, but has its principal limitation in the free satellite resolution. For most of the available imagery sources, an entire field is represented by few pixels, making the remote monitoring process impossible. The development of resolution enhancement algorithms can be vital for this research field and can open the way to new commercial solutions for smart agriculture.
This thesis studies the problem of super-resolution, that is the process of synthetically increase the resolution of an image, trying to recreate additional pixels as coherently as possible with the original information. The approach adopted to assess this problem is the development of a custom convolutional neural network (CNN), able to generate a high-resolution (HR) version of multiple available low-resolution (LR) images.

2. Proba-V challenge

The work has been developed with the primary objective of taking part to the PROBA-V Super Resolution challenge, organized by the Advanced Concept Team of the European Space Agency. The competition provides a dataset of pictures took by the PROBA-V satellite of the ESA that consists in 1160 training scenes of the Earth surface, divided into RED and NIR spectral bands, for which one HR image and several LR images are provided. The goal of the challenge is to provide the HR images for 290 testing scenes for which only the LR images are available. LR images has 128×128 pixels, while HR 384×384, thus the upscaling factor is 3. In addition to these images, quality maps are also provided for each LR and HR image: these binary maps mark with a 1 reliable pixels and with a 0 concealed pixels (usually due to cloud coverage).

3. Proposed model

Super-resolution algorithms are divided in two main categories:

- Single-image Super-resolution (SISR): tries to reconstruct a HR image from a single LR image.
- Multi-image Super-resolution (MISR): uses several images of the same scene and is based on data fusion techniques; the basic idea behind MISR is to exploit the non-redundant information that comes from the subpixel shifts between the different LR images.

Since the dataset provided for the challenge is composed of different scenes with several LR images, a MISR algorithm has been studied.

The proposed architecture is divided in two parts:

(i) the first block performs a SISR upscaling of 9 input images independently

(ii) the second block merges together the upscaled images to reconstruct a single HR output

The 9 input images are selected from the dataset with a threshold on the number of clear pixels (unconcealed from clouds) present. To do so, a simple algorithm that compares the quality maps of the LR images to that of the correspondent HR image is used.

The first part of the model is called Multi SISR, since it performs single-image super-resolution of multiple images independently. EDSR (enhanced deep residual network, proposed by Lim *et al.* in 2017) has been chosen as architecture, since it represents the SISR state of the art. Since the main objective of this block is to enhance the resolution of an image, independently from the others, the block is trained using as inputs a downscaled version of the HR images, instead of the LR images provided in the dataset, that can have sub-pixel shifts or different cloud coverage. After the training process, the LR images are fed to the network to obtain their upscaled version.

The second part of the model, called MergeNet, has the aim of merging the 9 super-resolved images in order to further improve the super-resolution process. This block has basically to perform three main operations and each of them is associated to a specific sub-network:

- extract cross-features between the various input images: Dense3D sub-network
- aggregate and reduce the extracted features: ReduceNet sub-network
- fuse the results to get a single output image: FusionNet sub-network

The first block is mainly composed of 3D convolutions; features extracted in a certain level are concatenated with features coming from previous convolutions, creating the so called densely connected architecture. The second block has the task of aggregating and reducing the information to get a more compressed representation and it is again based on 3D convolutions, since it manipulates blocks of cross-features extracted by the 9 input images. The last sub-network fuses the compacted cross-features together and outputs the super-resolved image, thus is based on 2D convolutions.

The MergeNet is trained end-to-end with the images obtained from the first block. Data augmentation is used, applying rotations and flips, to artificially increase the number of available input images. A schematic representation of the model overall architecture is shown in figure 1.

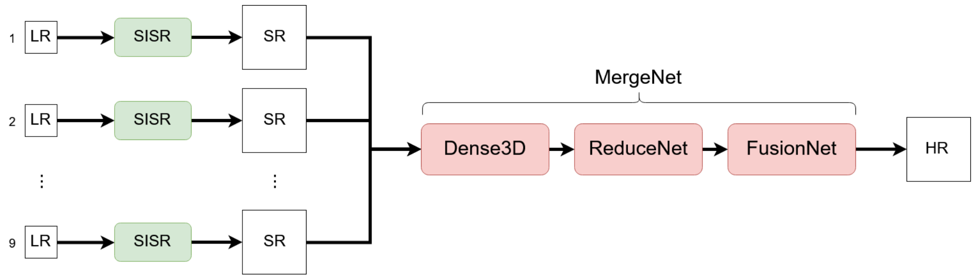
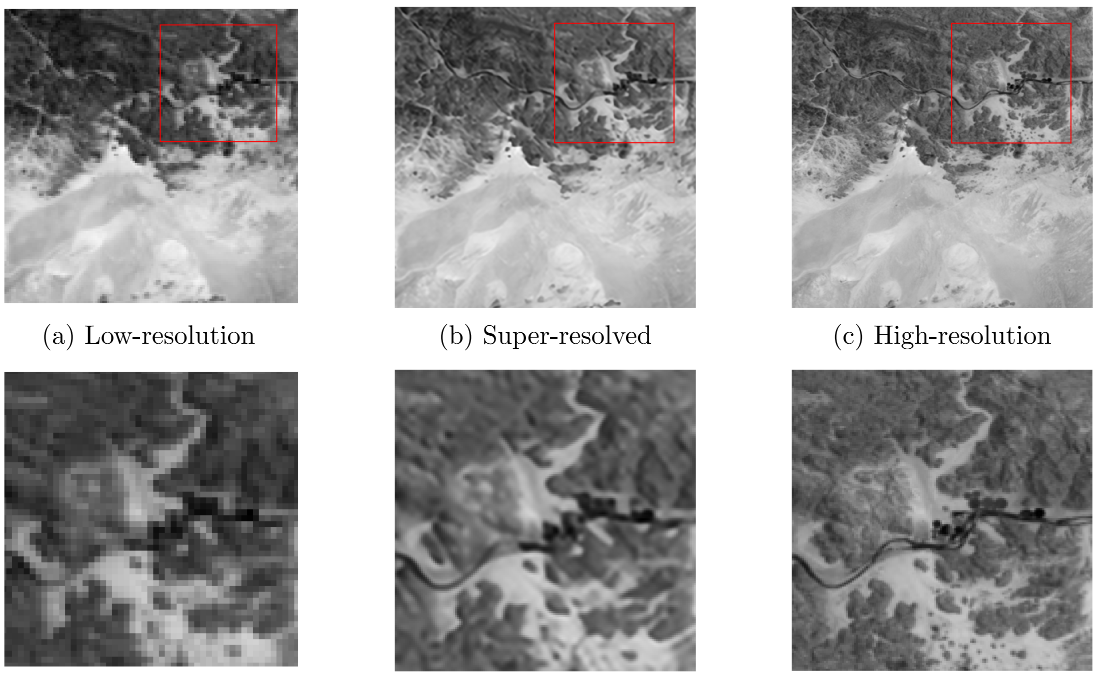


Figure 1: Model architecture

4. Achieved results

To understand the performance of the model, the bicubic interpolation method, that is classically used to upscale images, is used as reference. The images generated with the proposed model outperform the bicubic for 99.3% of the scenes. The test images have been super-resolved and submitted and reached the fifth place in the challenge final leaderbord. An example comparison between the low-resolution input, the super-resolved output and the high resolution target is shown in figure 2.
The model obtains better results also with respect to the state of the art single-image super-resolution architecture, showing that using multi images instead of a single one, can boost the super-resolution process, allowing to further increase the quality of the output images. This demonstrates that deep learning can be successfully applied to super-resolution, especially for satellite imagery, that is characterized by frequent reshots of the same scenes.
Further work can be perform to try to further improve the model, especially for what concerns efficiency and hardware resource demand. An interesting application of the algorithm can also be the satellite/drone imagery mapping, that consists in trying to super-resolve free satellite images, such as those of the Sentinel-2, using drone images as HR reference.



(a) Low-resolution (b) Super-resolved (c) High-resolution

Figure 2: Comparison between low-resolution, super-resolved and high-resolution images.