Spatial resolution is one of the basic properties of digital images, commonly expressed by the number of pixels in horizontal and vertical dimensions. It is worth noting that the number of pixels in the image alone does not determine whether sufficiently small details will be visible in the image, because other factors related to the quality of visual information play an important role as well. Certainly, we will not be able to recognize objects whose dimensions are smaller than a single pixel in the image, but it is not said that it will be possible in the case of objects covering larger areas. For this reason, increasing the pixel density is usually a necessary, while not sufficient condition to provide better spatial resolution, understood as the ability to distinguish sufficiently small details.

Currently, almost everyone knows a computer is capable of "enlarging" digital photos and images that are too small. This produces a larger image indeed (i.e. composed of more pixels), but usually we cannot see much more there than in the small version we had at the beginning—the image is hazy, blurred, and "pixelated". Such outcomes are obtained by the means of interpolation, which estimates the colour of new pixels based on the "old" pixels from the original small image. Most probably, a human would approach such a task in a more smart way. For example, if the small input image allows us to recognise an object whose appearance we known (e.g. a speed limit road sign), we could reproduce its appearance precisely in high resolution. In this regard, we would use our knowledge that the sign is round-shaped, has a thick red border and a number in the middle. In addition, using our artistic and graphic skills, we would be able to faithfully reproduce such a sign in the enlarged image using professional or amateur software. One can imagine that the number visible in the input image would be too small to be recognised—in such a case we may still try to guess it, so that the possible value appears in the enlarged image (for example, we would not enter the number '58', because we know that it is much more probable that '50' is displayed there). Importantly, the risk of committing a mistake would be lower, if we were presented not just a single small image, but multiple images, differing from each other, each of which would carry a different part of high-resolution information.

So-called *super-resolution reconstruction* algorithms, especially those based on machine learning, mimic the approach described above—they magnify a single input image or a series of images using knowledge of how individual objects look at low and high resolution. They acquire this knowledge on the basis of a properly constructed training set, containing paired images presenting the same content at different resolution. In recent years, we have witnessed a dramatic progress in super-resolution reconstruction methods, which is attributed to the use of convolutional neural networks in conjunction with deep learning techniques. The super-resolution capabilities have definitely increased, but they are still insufficient for real-life applications. Yes, in many cases they allow us to achieve amazing results, especially when applied to images that have previously been artificially degraded, but their behaviour is not stable and reliable enough to be used in an uncontrolled environment. This is particularly the case when critical decisions are to be made on the basis of the reconstructed information, for example in medical imaging or remote sensing, especially if some unusual data are presented for reconstruction. For example, if a medical image presents a pathologically changed tissue instead of the healthy one, it would be extremely dangerous, if the reconstructed image showed a healthy tissue, thus losing the information critical to the proper diagnosis. Referring back to the previous example, sometimes we have to be ready for the number '58' to appear on a road sign, even though there is no such officially registered road sign, and we do not want super-resolution to be guessing too much.

The goal of this project is to develop new super-resolution reconstruction methods with enhanced reliability that can be used in an uncontrolled environment for real-world images (i.e. not only those that have been artificially degraded). This will be achieved by applying a new strategy of training the deep networks that perform the reconstruction, which will consist in guiding the network, so that the reconstruction result is suitable for numerous advanced computer vision tasks. These tasks will include text recognition, detecting small objects, or segmenting some specific image regions (e.g. forests in satellite images or human skin in colour images). In addition, we will strive to develop self-supervised learning techniques that will allow the networks to be trained from unpaired low- and high-resolution images (i.e. which do not show the same scene). The requirement of using paired images (that is, presenting exactly the same scene) is very difficult to ensure and makes it challenging to collect sufficient amounts of real-world data. Therefore, lifting or relaxing this limitation will facilitate the development of super-resolution reconstruction methods suitable for real-world applications.