

Image Super Resolution Using Generative Adversarial Networks and non-Paired Strategy

Letícia Karolina Moreira¹, Marcelo Romero², Manassés Ribeiro¹

¹Catarinense Federal Institute of Education, Science and Technology, Videira, Brazil

²Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

leticia.moreira@yandex.com, nmarceloromero@gmail.com, manasses.ribeiro@ifc.edu.br

Abstract—The quality of images obtained from video surveillance systems is a decisive aspect when performing investigations at the Forensic Science. Features such as scars, tattoos, and skin marks are great examples of details that allow to consolidate an investigation at certain scenarios in which there is the necessity to identify individuals captured in a video or image footage. However, the low quality of images could affect the results of the investigations. In this sense, this work proposes the study of a computational model to address the problem of increasing the resolution of Low-Resolution (LR) images, also known as the problem of super-resolution of images. The main idea is to train a Generative Adversarial Network (GAN) so that it can be able to enhance low-quality images. The hypothesis is that a variant model of a GAN, named Super-Resolution Generative Adversarial Network (SRGAN), is capable to produce High-Resolution (HR) images from LR ones. The proposed methodology is based on experimental research with the aid of the hypothetical deductive method, where two well-recognised state of art methods were used, which proposes the use of convolutional neural networks and deep learning. For the model validation, were conducted four different experiments: two to avail the capacity of the GAN to produce images with enhanced resolution and two other experiments to evaluate the quality of the results produced by the SRGAN. The quantitative results of our experiments are promising, with performances that are similar to those obtained by state-of-the-art approaches. Moreover, the qualitative results based on performing a visual analysis of the images produced by our approach suggest a interesting performance in terms of visual quality.

Index Terms—Super-Resolution of Images; Generative Adversarial Network; Convolutional Neural Networks; Deep Learning.

I. INTRODUCTION

In recent decades, technological advances, both in hardware and in software, have allowed significant improvements in the quality of digital images, from low-quality images to images of comparable quality to photographic ones. This is currently observed in most smartphones and digital image capture equipment. In this sense, an area that accompanies such development, although at a slower pace, is the field of surveillance camera systems.

In spite the presence of surveillance systems available in high definition resolutions (full HD), factors such as their purchase price and maintenance cost often makes it difficult to access this equipment for use of families, small businesses and places where there is a need of a high number of monitoring equipment. Another reason that often influences the decision of which image resolution to use in surveillance equipment is

the amount of storage capacity that is required for the videos and images. These devices, in general, need to store images for long periods and for many days, which ends up limiting the space that can be used for recordings. Low-quality captures take up considerably less space, which makes it an important factor to consider when choosing which quality level will be employed for surveillance systems, generally resulting in choosing the path of the use of low-resolution images.

In the area of forensics and criminal investigations, one of the most important resources that can influence (both positively and negatively) an investigation, are visual resources used for such purpose, including photographs and footage from surveillance cameras. Graphic details such as tattoos or unusual marks can help during the investigative work for the necessary clarifications.

Due to the sensitivity required for the correct analysis and interpretation of images, low-resolution captures often end up affecting forensic science and its sub-areas. This issue could even hinder the progress and results of investigations that could be promising. According to [1], footage of robberies and kidnappings released by the police is usually so blurry and *pixelated*¹ that specific details are practically non-existent. In [1] it is also mentioned that better monitoring video qualities would make investigations much more effective and sharper. In the Brazilian scenario, difficulties persist and are often even more aggravating, as demonstrated by [2], where the Bahia state police say that many cameras used on bus lines are of such poor quality that they do not allow for identification of criminals. Although there are techniques and tools that allow to improve image quality, most are manually-based, which means that a human must insert one image at a time in the tool to obtain results, which could become an expensive task in terms of time and resources. Furthermore, most of these methods are based on the use of pre-arranged High-Resolution (HR) images to make a pairing and mapping with the inserted Low-Resolution (LR) images. Therefore, if the input image does not have a similar pair in the database that is used for comparison, the results generated do not show significant improvements in relation to the input images.

Taking these aspects into account, it is common that many of the recently proposed methodologies, such as Generative Adversarial Network (GAN) for super image resolution [3],

¹Clear visualisation of large squares (pixels) in a digital image.

[4], try to circumvent the problems specified before. However, they end up using approaches in which the pairing between HR and LR images is necessary. Therefore, they are not directly applicable to real-world problems, especially when it comes to surveillance camera systems, since the images in their counterpart² are non-existent, as in the aforementioned approaches. Some proposals [5], [6] seek to increase the resolution of LR images without the existence of its counterpart, making it a more viable method for real-world problems.

This work aims at targeting the issue described before through the use of GANs to increase image quality (super resolution) in surveillance camera systems. The main working hypothesis is that, with a computational model that uses GANs, it is possible to reconstruct a LR image into an HR image even without having a counterpart image available. In this sense, the main objective of this work is to propose a computational model for the improvement of LR images using GANs, which can serve as a tool to aid at performing computer forensics tasks. The following specific objectives are derived from the initial main objective:

- to propose a methodology for improving forensic images through the use of GANs;
- to assess the ability of a GANs to generate synthetic HR images from LR images;
- to verify the ability of a GANs to generate HR images from known LR images;
- to quantitatively and qualitatively analyse the quality of HR images generated by a trained GAN model;
- to evaluate the computational model designed using publicly available datasets.

This paper is organised as follows: the related work is presented in Section II. The Methodology, which describes step-by-step the strategy to overcome the problem, in addition to the GAN architecture, data preparation and model evaluation metrics, are presented in Section III. The computational experiments, as well as their results and discussions are presented in Section IV. Finally, Section V presents the conclusions of this work and possible future research directions.

II. RELATED WORK

In general, recent works in literature that propose to overcome the image super resolution problem use a paired approach, where the idea is to use a pair of HR and LR images, whose LR image is the counterpart of the HR image. For works that use this premise, HR images are interpreted as being true samples and are used as input to the Discriminative Network (DN), while LR images are used as input to the Generative Network (GN) in order to produce HR images that could be accepted by DN. Examples of works that use this methodology are proposed by Ledig *et al.* [3], Li *et al.* [4] and Wang *et al.* [7].

Ledig *et al.* uses a GAN variant where LR images have their resolution increased by means of up-scaling convolution

layers. The model is composed of both GN and DN networks, where GN is trained with the LR images to produce HR images, which are confronted with the counterpart by the RD model. The reported results show that the approach outperforms reference methods by a large margin and defines a new state-of-the-art for super resolution at photo-realistic images. In this way, Li *et al.* proposes a work using the original architecture proposed by [3], to be applied to the problem of reconstructing textile images, and to the failure recognition. Here authors present a comparison of GAN-based approach with two other image enhancement techniques, *i.e.* bi-linear and CNN-based. Results suggest that both GAN and CNN approaches stand out when compared to the bi-linear method. Besides, GAN-based method is able to present visually more detailed image, while the image obtained by CNN model is smoother. As evaluation metrics, authors suggest the use of both Structural Similarity (SSIM) [8] and Peak Signal to Noise Ratio (PSNR) [9] due to their capability to compute the distance between the real and generated images, thus providing quantitative results that may be used as parameters in a visual comparison. Wang *et al.*, in turn, propose an “enhanced” version of the GAN-based approach proposed by [3]. The work idea is to propose improvement strategies to the network architecture and also to optimise both the model loss function and the adversarial function, with the objective of improving the overall performance of the model.

More recently, Ren *et al.* [10] proposes a method inspired by [7] to circumvent the image SR problems, and Alam *et al.* [11] propose an SRGAN approach, supported by model introduced by [3], for imaging microscopy. On the other hand, Lin *et al.* [12] uses a SRGAN approach to circumvent the previous problem regarding multiple degradation on GANs, where the method idea is to use three discriminators, at the RD training, for improving the accuracy. The obtained results show that the proposed model deals well with multiple degradation and produce images with wealthy features.

Important mentioning that aforementioned works use paired approach. However, these models are not suitable for real-world problems, since images in real-world contexts do not have a counterpart in HR and, if they did, there would be no purposes to perform artificial images in HR. Thus, to circumvent this problem Lian *et al.* [5] propose a ‘feature-guided’ approach, which uses as input to the DN images previously generated by a GAN. The problem here is that the images used for training the GAN are part of a pre-defined HR image dataset, which can limiting the work scope. The interesting aspect of this approach is the use of the “feature-guided” module, which increases the model performance.

Notwithstanding, Shocher *et al.* [6] propose an unpaired GAN-based model to be used in super resolution image problems (SRGAN). Although results obtained by this approach are slightly inferior compared to other paired GAN-based approaches, this model becomes more attractive due to the perspective of using it at real-world problems, where there are no counterpart images in HR. To circumvent the counterpart problem, authors propose the use of Zero-Shot

²Counterpart refers to the opposite plane of the mentioned image. If an image is in HR, its counterpart is the same image in LR and vice-versa.

strategy, which is composed of a CNN trained specifically for each input image, thus allowing the approach to adapt to different application contexts. The CNN is trained to learn the complex relationships between both HR and LR domains, which are then applied to the input LR image in order to produce an output HR image. It is important mentioning that, despite the training cost, the results produced are similar to those obtained by state-of-the-art approaches that use synthetic LR datasets and surpass those whose counterpart in HR is non-existent.

In short, the idea of this work is to use approaches proposed by [6] and [5], with respect to the unpaired GAN-based super resolution images and the use of “features-guided”, respectively, as a starting point to propose an unpaired image enhancement model (image up-scaling) that can be used in real-world problems, *e.g* images from surveillance cameras.

III. METHODOLOGY

This section presents methods to circumvent the problem introduced in this work, which refers to the use of a GAN for increasing the quality of images (super resolution) from surveillance camera systems. To better understand the proposed method, this section is divided into subsections as follows.

A. Model Overview

The suggested method is composed of three main steps, which are *Data Preparation*, *Training and Tests*, and *Evaluation and Analysis*, as illustrated on Figure 1. The *Data preparation* step is concerned with how the data needs to be prepared and pre-processed for the next steps. At the *Training and testing* stage, the GAN architecture, training procedures and the test protocol are defined. Finally, at *Evaluated and analysis* step, metrics that will be used to assess the model performance are described.

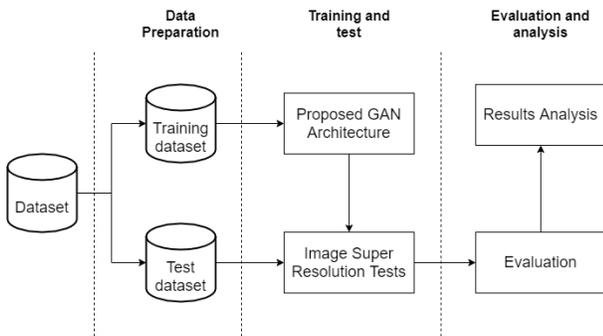


Fig. 1. Overview of the proposed model.

B. Data Preparation

First, images need to be resized to the 32×32 size, since it is the standard input resolution to the GAN. After, the dataset need to be split into training and test subsets, which for this work we choose a ratio of 70% for training and the remaining 30% to test.

C. Model Details

This section presents all necessary steps to generate High-Resolution (HR) images from Low-Resolution (LR) images. In short, the model is composed of Zero-Shot Super Resolution (ZSSR), Discriminative Network (DN), Generative Network (GN) and a Mediator Module (MM). ZSSR will be responsible for producing synthetic HR images, whose will be fed as training samples to the DN. On other hand, original images (in LR) will be fed to the GN and at MM is computed the loss function, thus extracting features from images learned from both DN and GN. The training process is achieved by minimising the loss between DN and GN outputs.

1) *GAN architecture*: The GAN architecture used in this work is based on the works [6] and [5]. From the first one we use the idea of generating HR images without the need of using a previous dataset that may be used as synthetic counterpart of the LR images. From the other one, the proposal of using the intermediary module (MM) to minimise the space of possible mapping functions will be adopted, allowing the GAN to learn a intrinsic mapping function from LR to HR domain.

In order to better understand the GAN architecture for HR purposes, the design of the network is shown at Figure 2. First, LR images are forwarded to the ZSSR module, whose result will be ‘synthetic’ HR images, to suppress the needed counterpart samples. With the synthetic dataset, the DN is trained at the same time that the original images in LR are fed to the GN, thus carrying out the GN training. The trend is that during the training progress, the GN learns how to generate better images, i.e learns images more similar with synthetic HR images, and then in fact one HR image is produced from another in LR. Finally, the MM is a VGG network used for extracting features from both GN and DN outputs, in order to limit the space of learned features, thus increasing the GAN performance.

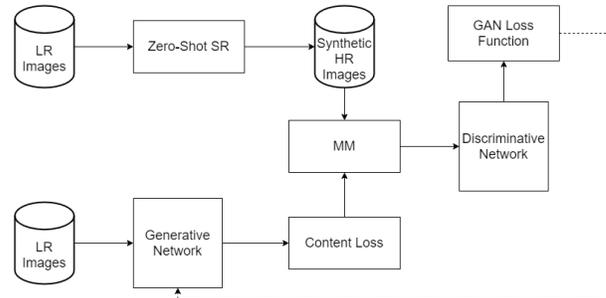


Fig. 2. Scheme of the Generative Adversarial Network for Super-Resolution in images.

2) *Zero-Shot Super Resolution*: The ZSSR is composed of a small Convolutional Neural Network (CNN), which is trained on sub-samples produced from LR images [6]. Given an input image I , a specific (tailored) CNN is built to perform the super resolution for that I image. The sub-samples are obtained by reducing the scale of the image I , thus generating a new version with LR, denoted by $I \downarrow s$, where s is the scale parameter. When the CNN training process is finished, the LR image I (original size) is fed to the CNN in order

to produce its HR counterpart, which is formally denoted as $I \uparrow s$. An illustration of how ZSSR is performed may be seen in Figure 3.

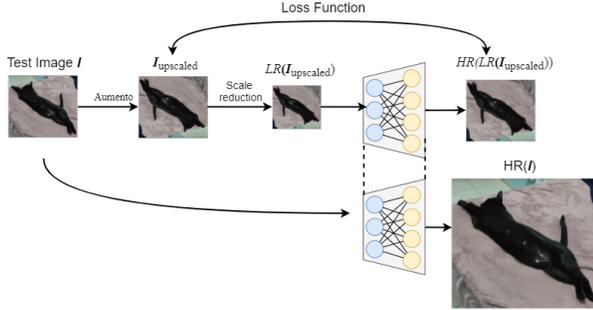


Fig. 3. An Illustrative Model of the ZSSR.

During the training process, a data-augmentation strategy is performed with purposes of increasing samples, since initial image is just one. This step is computed with an alternation of four different rotations, 0, 90, 180 and 270, respectively, and the use of mirror reflection in both vertical and horizontal rotation. Due to these procedures, each I LR image resulting in another 8 samples in HR for training step, thus tending to produce better results, once the model has more generalisation capability.

3) *Discriminative and Generative Networks*: The DN architecture is composed of an input layer, 8 blocks with 3 layers each (Conv2D, LeakyRelU and BatchNormalisation), which are related with the feature extraction step, and both flatten and dense layers that composes the classifier. Regarding the GN architecture, it is composed of a amount of 86 layers: 1×Conv2D layer, 1×PReLU layer, 16×Residual blocks, 1×Conv2D layer, 1×BatchNormalisation layer, 2×Up-scaling blocks, and 1×Conv2D layer. The Residual block, in turn, is composed of 5 layers, i.e., 1×Conv2D layer, 1×BatchNormalisation layer, 1×PReLU layer, 1×Conv2D layer, and 1×BatchNormalisation layer. It is worth mentioning that at Residual blocks there is a skip connection strategy, where layers are skipped randomly every epoch training. The idea of this scheme is to avoid the network saturation problem. On the other hand, the Up-scaling block is composed of 1×Conv2D layer, 1×UpSampling2D layer, and 1×PReLU layer. Notice that the UpSampling2D is a layer that aims to double dimensions from an input 2D data. Details of both DN and GN architectures can be seen at the work of Ledig *et al.* [3].

4) *Mediator Module*: The MM consists of a VGG-19 Network, which is a variant of the VGG³ model, and is composed of 19 layers (16 convolution layers, 3 Fully connected layer, 5 MaxPool layers and 1 SoftMax layer) [13]. VGG-19 was pre-trained on ImageNet, which is a dataset that encompasses over 14 million of images from various categories.

³VGG is a kind of a CNN. More detail can be found at <https://neurohive.io/en/popular-networks/vgg16/>.

D. Training Stage

The training process of both DN and GN occurs concurrently during a certain number of epochs, using parameters that will be addressed in Subsection IV-C. First, the DN is trained for some epochs and then GN is also trained for some epochs. The objective of the DN training is to make this network able to learn to identify real data, besides learning how to recognise possible failures produced by GN. Note that since the training process is concurrent, while the first one architecture is being trained (e.g DN), the last one remains with the fixed weights (e.g GN), alternating after a number of epochs is done. This alternating process is repeated until the end of the training.

E. Test and Evaluation Stage

After the GAN training is completed, the model testing is carried out. The test is performed only with the samples from the test subset as described in III-B. Next, the Evaluation and Analysis step is performed, for which two metrics are used: SSIM and PSNR. The SSIM measures the structural similarity between two samples, while the PSNR calculates the signal noise peak from the two samples. Although there are other metrics that may be used to evaluate the GAN results, we chose these for their capability to compute the distance between the real and generated image, thus providing quantitative results that we can use as comparison parameters for visual analysis of the improvement of GAN image in HR compared to the original LR image.

IV. EXPERIMENTS AND RESULTS

This section describes the technologies used to implement the proposed approach, the datasets used for the experiments, the parameter settings for the training and testing phases of the model, and the computational experiments that were proposed to verify the working hypothesis introduced in Section I.

A. Technologies

The implementations presented in this section were developed using Python [14], version 2.7. Although there exists a newer version of the language (3.7.x), an older version was chosen to be compatible with the ZSSR, which was developed for Python 2.7 only. We also used the TensorFlow library, version 2.0, together with Keras [16]. The choice was made due to the possibility of Keras to act as an interface for several machine learning *frameworks*, including TensorfFlow, and due to its ease of use and its capacity to be extended.

For both experiments IV-D and IV-E, a General Purpose Graphic Processing Unit (GP-GPU) was used, given the fact that the work encompasses the use of a dataset with a large amount of images, where the number of calculations and processing required are high. The GP-GPU used in this work was the Nvidia Geforce Titan XP, with 12 GB of RAM, and performance of up to 12 Teraflops. On the other hand, for experiments IV-F and IV-G, Google Collaborator was used, a free code execution environment in the cloud. Each environment provides 12GB of RAM, 50GB of cloud storage, and available GP-GPUs for resource-demanding processing.

Google Collaborator was selected since it is easy to use and does not require any additional configurations.

B. Dataset

For this work, two publicly available datasets have been used, entitled QMUL-TinyFace⁴ [17] and QMUL-SurvFace⁵ [18]. The *Tinyface* set consists of 169,403 native images in LR containing faces of people, with an average resolution of 20×16 pixels per image. The images were captured in a large variety of scenarios, with different variables of pose, occlusion, angles, lighting and background. The *Survface* set consists of 463,507 original LR images, also containing faces of people, captured in non-cooperative (non-ideal) scenarios of real-world surveillance camera systems. Both sets share characteristics that are common for footage obtained from surveillance camera systems: their images have low quality and very low resolution.

Since there are two architectures that make up the GAN (DN and GN) two subsets of data were used. For the DN, the training set consists of images sent to the ZSSR for resolution enhancement, interpreted as real data. The remaining images are directed to the GN, which generates synthetic samples that are then forwarded to the DN to be tested. Since the GN uses the same dataset for its training phase and to later test the DR, a training-test dataset split is not necessary in this scenario.

C. Training Settings

1) *Zero-Shot Super Resolution*: As parameter of the ZSSR, the scaling rate was set to be equal to 2×2 , this implies that every image inserted in the network has its resolution doubled. For instance, given an image with an initial resolution of 26×29 pixels, its final resolution in HR will be 52×58 pixels.

For the learning rate⁶, the initial value suggested by [6] was used for our experiments: 0.001. Periodically, a linear adjustment of the reconstruction error is performed, and if the standard deviation value is greater than the slope of the linear adjustment, the learning rate is divided by 10. The stopping criterion is defined according to the learning rate, defined for when its value reaches to 10^{-6} .

Adam is used as optimisation function⁷, which was originally introduced by [19]. The number of epochs has an undefined value, given the fact that the ZSSR training stopping criterion is based on stopping the training when a certain value for the learning rate is reached.

2) *Super Resolution Generative Adversarial Network*: The resolution increase rate defined for the SRGAN is the same as defined for the ZSSR, that is, a scale of 2×2 pixels. Likewise, all images inserted in the GN have their final resolutions doubled. The number of epochs defined for both the DN and

GN training phases was equal to 2,000, as suggested in the work by [20].

The value that was defined for the *Batch Size*, which is the number of examples used in each iteration during the training phase, is equal to 50. Although a high-performance GP-GPU often allows for higher *batch Size* values, other factors must be considered when dealing with a GAN. As pointed out by [21] and [20], a value of *batch Size* that is too large for a GAN could significantly increase training time and affect the overall performance of the model, potentially causing a decrease of its performance. Therefore, small values are recommended to be used so as to obtain better performance of the network.

Same as for the ZSSR is used, the Adam optimiser was chosen to train the model. As for the loss function, we used *Binary Cross-entropy*. This function was chosen since the SRGAN has only two possible classes: an image can belong or not to the synthetic HR set.

D. Experiment 1: HR Images Generation using ZSSR

The first experiment was based on producing a set of synthetic HR images from the ZSSR. The objective of this experiment was to verify if it is possible to use the ZSSR to reproduce images in HR, which can be used to train the DN as a synthetic dataset, in order to suppress the lack of an original HR images required to train the SRGAN.

After data preparation (see Section III-B), a group of 14,000 native images in LR were fed to the ZSSR. Although the datasets add up to a total value of 632,910 samples. The number of instances has been reduced so as to reduce the time needed for the ZSSR to produce all the images. Moreover, this allows to not compromise the performance of the SRGAN in the training phase. Figure 4 illustrates some samples used as input to the ZSSR.

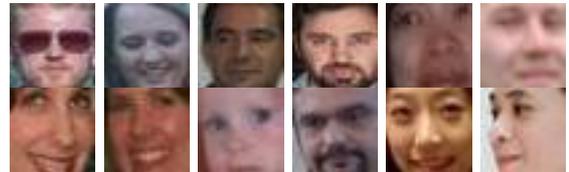


Fig. 4. Examples of images that were fed to the ZSSR to generate the synthetic dataset.

The training/test stage of the ZSSR was divided into two distinct procedures. First, a set of 7,000 images was used as input to the model, coming from the QMUL-Tinyface dataset. This procedure took about 120h for the network to be trained and also to test the entire dataset. Due to the fact that the ZSSR tests its inputs at the same time it trains, it was possible to make the individual visualisation of each result as they were being tested.

The second step was to use 7,000 samples from the QMUL-Survface dataset. As a result, 14,000 images were produced in HR, with their respective final resolutions doubled, being 64×64 pixels each. This set produced later is forwarded to the DN. Figure 5 demonstrates some of the results produced by ZSSR.

⁴Available at: <https://qmul-tinyface.github.io>

⁵Available at: <https://qmul-survface.github.io>

⁶A fitting rule in an optimisation algorithms that determines the step size in each iteration, while moving to minimise the function of model loss.

⁷Responsible for adjusting the weight parameters to minimise the loss function of the model.



Fig. 5. Samples of the dataset composed of synthetic images produced by the ZSSR.

Figure 5 shows that, in addition to the increase in resolution, the characteristics of the images remain more visibly remarkable. Moreover, it is also possible to observe more clearly the details of the images. The result was satisfactory, thus the generated images were used as a synthetic dataset of HR images to train the SRGAN.

E. Experiment 2: HR Images Generation using SRGAN

Once the set of synthetic HR images were generated, the second experiment took place. This experiment consisted of training and test the SRGAN. The objective of this experiment was to study the capability of a SRGAN to produce HR images from real-world low quality images.

All images from the synthetic HR dataset were fed to the DN, and a set of 6,000 samples from the initial LR dataset were fed to the GN. The main idea was to train the GN to produce better and better images with each epoch, in order to achieve the capability to 'deceive' the DN so that its results could be accepted as an original HR image by the DN. Figure 6 demonstrates some sample images that were used as input for this experiment.

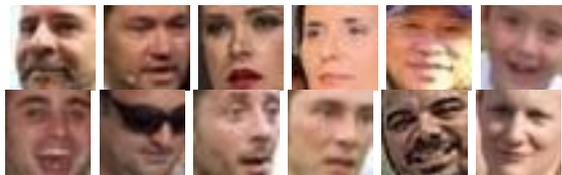


Fig. 6. Samples of the dataset used to train the SRGAN.

After the SRGAN training, 6,000 HR images were obtained, each with a resolution equal to 64×64 pixels. Some examples of the results produced by SRGAN are shown in Figure 7.



Fig. 7. Examples of HR images produced by the SRGAN using the input images shown in Figure 6.

Figure 7 shows that some details from the original images were blurred. Notwithstanding, it is possible to visually notice a large improvement of the quality of the images. A more detailed analysis of these results is presented in sections IV-F and IV-G, where the results obtained by model are discussed from quantitative and qualitative perspectives, respectively.

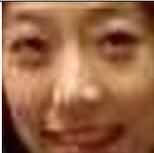
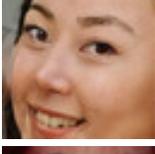
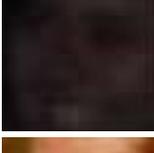
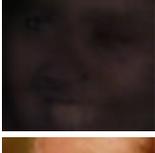
F. Experiment 3: Quantitative Analysis of the Results

This section presents an evaluation experiment that consists of using two image reconstruction evaluation metrics: SSIM and PSNR. The objective of this experiment is to infer if the results produced by SRGAN are optimistic. For this purpose, images in LR are used, which also compose the set of images inserted in the ZSSR. This is done because, for these measurements to be used, there must be a real image (R), in HR, and a generated image (G), which was produced by SRGAN.

The SSIM-PIL [22] is implemented to calculate the SSIM. In this approach, two images of the same resolution are compared, and as a result, a value between 0 and 1 is obtained. A value close to 0 means that the images are completely distinct and a result close to 1 implies that the images are identical. On the other hand, the PSNR is implemented through a function of Scikit-learn [23], where the real image (original), the test image (generated) and the minimum and maximum possible range of the result are passed as parameters for the function. As a result, the PSNR between the two images is returned in decibels (db s).

Table I illustrates a comparison of the images of the set in synthetic HR and the images in HR generated by SRGAN considering the metrics mentioned before.

TABLE I
COMPARISON BETWEEN THE SYNTHETIC HR IMAGES AND THE HR
IMAGES GENERATED BY THE SRGAN.

ID	Synthetic Image	SRGAN Image	SSIM	PSNR
1			0,9132	31,5044
2			0,8568	31,6848
3			0,8951	31,9378
4			0,8668	31,2329
5			0,8566	31,1281
6			0,8715	31,2964
7			0,9462	38,9739
8			0,8835	32,9032

As shown in Table I, for the SSIM index, the results obtained have similar values, ranging between 0.85 and 0.89, with emphasis on the sample ID 7, which obtained a SSIM the value of 0.94, very close to 1. In the context of the SSIM metric, this value suggests that the images are very similar. On the other hand, according to [24], the typical values for noisy or small images are between 30 and 50 dB when the PSNR metric is considered. Therefore, it can be considered that a satisfactory result has been achieved for all tested images.

It worth mentioning that a higher SSIM or PSNR index for

a certain images does not necessarily indicate that, in terms of visual quality, it was the best generated image. This is possible to verify by inspecting the image with ID equal to 7 in Table I, which achieved SSIM and PSNR equal to 0.89 and 38.9 db, respectively. The image does not seem to have best quality, if a visual inspection is performed. Although its resolution has been increased, some of the image details are blurry, making it difficult to see.

G. Experiment 4: Qualitative Analysis of the Results

The fourth and last experiment was based on performing a visual analysis of a set of HR images produced by the SRGAN, whose counterparts in LR were not fed to the ZSSR. The aim of this experiment is to investigate the capability of the SRGAN to produce good HR images (from a visual perspective). As these images do not have original counterparts in HR, it is not possible to use metrics like those implemented in the Experiment IV-F.

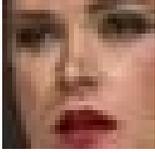
Table II presents a comparison between images in synthetic HR, extended versions of these (using *zoom*) and finally the HR image results produced by the SRGAN.

As shown in Table II, the HR images produced by SRGAN achieved a considerable improvement in terms of visual quality, level of detail, and resolution. It is important to emphasise that in this case it is not possible to perform a quantitative analysis using the image reconstruction evaluation metrics used in the previous section, since there is no set of original HR images for comparison.

V. CONCLUSION

This paper introduced a method to transform Low-Resolution (LR) images into High-Resolution (HR) ones. This is an issue that is specifically relevant to aid during the analysis of footage extracted from surveillance camera systems. The hypothesis of our work states that the use of a Generative Adversarial Network (GAN) allows to reconstruct a LR image into an HR image, even without having a counterpart image available. The architecture proposed in this work was inspired by two previous works. The first one, referred to here as ZSSR, was used to generate synthetic images in HR. The second one, called Super-Resolution Generative Adversarial Network (SRGAN), inspired the use of an intermediate module to reduce the space of possible mapping functions. Specifically, the VGG-19 network was used pre-trained in this work. For quantitative assessment of the results, two image reconstruction functions were used: Structural Similarity (SSIM), and Peak Signal to Noise Ratio (PSNR). Moreover, a qualitative assessment based on visual inspection was also performed. The results obtained in this work suggest that the proposed approach is promising and it can be applied in contexts where the generation of high-resolution images is necessary, such as forensic areas. This approach suggests that a GAN model can be useful as an aid tool in analysing poor resolution images, such as those that are found in most surveillance camera systems.

TABLE II
COMPARISON BETWEEN ORIGINAL LR IMAGES AND HR ONES
GENERATED BY THE SRGAN.

LR	Extended LR	HR
		
		
		
		
		
		
		

Future work could aim at the development and application of a *software* that can deploy the SRGAN in an industrial scenario. Moreover, tests of the software tool with the aid of professionals in the Forensic area could be performed with the objective to validate how useful the tool can be for eventual forensic investigations. Likewise, the use of down-scaled versions of high-res images can be addressed to build a dataset that allows to check the quality of the super resolution versions of those down-scaled images. In addition, a comparative work can be proposed between the SRGAN proposed in this work and other methodologies that are considered as state-of-the-art, such as [3], [25] and [6].

ACKNOWLEDGEMENT

All authors thanks NVIDIA for the donation of Titan-Xp GPU boards used in this work.

REFERENCES

- [1] A. Carter. (2014) Why is security camera video still so terrible?
- [2] BusVision. (201-) Câmeras de baixa qualidade prejudicam o trabalho da polícia na bahia.
- [3] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016.
- [4] J. Li *et al.*, "Super resolution image reconstruction of textile based on srgan," in *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*. Tianjin, China: IEEE Computer Society, 2019, pp. 436–439.
- [5] S. Lian, H. Zhou, and Y. Sun, *FG-SRGAN: A Feature-Guided Super-Resolution Generative Adversarial Network for Unpaired Image Super-Resolution*. Cham, Switzerland: Springer, 06 2019, pp. 151–161.
- [6] A. Shocher, N. Cohen, and M. Irani, "'zero-shot' super-resolution using deep internal learning," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. EUA: IEEE Computer Society, June 2018.
- [7] X. Wang *et al.*, "ESRGAN: enhanced super-resolution generative adversarial networks," *CoRR*, vol. abs/1809.00219, 2018.
- [8] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ser. ICPR 10. EUA: IEEE Computer Society, 2010, p. 23662369.
- [9] A. Borji, "Pros and cons of gan evaluation measures," 2018.
- [10] H. Ren *et al.*, "Real-world super-resolution using generative adversarial networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 1760–1768.
- [11] M. S. Alam *et al.*, "Super-resolution enhancement method based on generative adversarial network for integral imaging microscopy," *Sensors*, vol. 21, no. 6, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/6/2164>
- [12] H. Lin *et al.*, "Generative adversarial image super-resolution network for multiple degradations," *IET Image Processing*, vol. 14, no. 17, pp. 4520–4527, 2020.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2014.
- [14] G. Van Rossum and F. L. Drake Jr, *Python reference manual*. Alemanha: Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [15] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015.
- [16] F. Chollet. (2015) keras. [Online]. Available: <https://github.com/fchollet/keras>
- [17] Z. Cheng, X. Zhu, and S. Gong, "Low-resolution face recognition," *CoRR*, vol. abs/1811.08965, 2018.
- [18] —, "Surveillance face recognition challenge," *arXiv preprint arXiv:1804.09691*, 2018.
- [19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization." *CoRR*, vol. abs/1412.6980, 2014.
- [20] T. Mittal. (2019) Tips on training your gans faster and achieve better results. EUA.
- [21] B. Ghosh *et al.*, "An empirical analysis of generative adversarial network training times with varying batch sizes," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*. EUA: IEEE Computer Society, 2020, pp. 0643–0648.
- [22] ChsHub. (2020) Ssim-pil. [Online]. Available: <https://pypi.org/project/SSIM-PIL/>
- [23] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [24] S. T. Welstead, *Fractal and Wavelet Image Compression Techniques*, 1st ed. EUA: Society of Photo-Optical Instrumentation Engineers (SPIE), 1999.
- [25] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," 2019.