A Convolutional Neural Network Model for Superresolution Enhancement of UAV Images

Daniel Gonzalez
AI Department
BQ
Las Rozas (Madrid), Spain
daniel.gonzalez@bq.com

Miguel A. Patricio
Applied Artificial Intelligence Group
Universidad Carlos III de Madrid
Colmenarejo (Madrid), Spain
mpatrici@inf.uc3m.es

A. Berlanga
Applied Artificial Intelligence Group
Universidad Carlos III de Madrid
Colmenarejo (Madrid), Spain
aberlan@ia.uc3m.es

José M. Molina
Applied Artificial Intelligence Group
Universidad Carlos III de Madrid
Colmenarejo (Madrid), Spain
molina@ia.uc3m.es

***Abstract*—In recent years, the use of unmanned air vehicles (UAVs) in various fields has become widespread. These UAVs have a set of sensors that allow obtaining information of the scenarios by which they fly, in order to monitor them or to be used in their navigation tasks. The camera is one of the most relevant elements. UAVs have limitations regarding their autonomy time. If you want to monitor a large geographical area in a short time, you need to make flights at a higher altitude. This implies loss of spatial resolution, since the cameras themselves have their limitations in terms of their optics and the size of the pixels. In recent years, super-resolution techniques based on deep Convolutional Neural Network (CNN) have been developed, which are able to learn the correspondence between low resolution images with their high-resolution counterparts. The problem of these models lies in the computation requirements that they need for their execution, not being viable to be executed in the embedded hardware of a UAV. In this work we propose a super-resolution method based on deep CNNs capable of being executed in the on-board equipment of an UAV.**

Keywords—Convolutional Neural Network, Superresolution images, UAV

I. INTRODUCTION

Unmanned air vehicles (UAVs) are evolving in the last decade to improve their versatility, flexibility, low-cost and minimized operational risk. Todays, UAV control is based on a set of complementary sensors (laser, sonar, Radar, imaging devices, etc.) [1-3] in conjunction with the classical GPS control [4]. This multisensory capacity improves UAV navigation although the integration requirements (consume, weight, dimensions) are much more restrictive in this kind of vehicles [5] than in Unmanned Ground Vehicles (UGV).

The use of UAVs has increased in various fields because this capacity to carry out a variety of sensors and, in this sense, UAV should be considered as a sensor platform able to acquire heterogenous data from the environment to define navigation and other functions. In this way, these sensor platforms have the potential to provide a huge variety of data on remote sites for applications as meteorology, sports, bushfire monitoring, detection and elimination of targets in defense, search and rescue, etc. [6].

One of the main useful technologies to understand the environmental situation is the video images. Actually, small device could be carried out in the UAV and real-time video processing could improve image information to improve real-time decisions. The use of onboard microprocessors and small cameras take advantage of real time processing to extract information in order to improve the next decision. Examples are:

- Video geo-localization [6] where a combination of low-cost GPS, video and attitude sensors are used to estimate the object's ground position.
- 3-D reconstruction from UAV images in agricultural problems [7]

Image acquisition from UAV has many benefits (aerial photos is a good way to analyse the situation) but have two big problems [8]:

- Quality of image from UAVs has many factors to be degraded: internal factors (sensor performance, exposure times, motion modes, etc.) and external factors (rain, fog, movement, illumination, etc.).
- Height should be maximizing to reduce amount of time to take images from large field

These problems affect to the resolution of the image. Extraction of information from low resolution images is not precise and this imprecision limits the extensive application of UAVs. In this sense, super-resolution enhancement techniques have been proposed [8].

The problem of superresolution (SR) is understood as one that seeks to obtain a higher resolution image from a sequence or a lower resolution image. SR has been applied to numerous real problems where it was necessary to obtain images with high spatial capacity. For example, the SR has been applied to medical images [9], remote sensing [10], surveillance [11], and in general in problems in which the SR facilitates the recognition of scenarios.

The spatial resolution in an image has a physical limit that is given by the size that a pixel can reach, so the solution has to be applying some technique based on software. In this sense, we find, on the one hand, solutions

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

based on the reconstruction of images from a sequence of images, and on the other, solutions that build the image of higher resolution starting from another of lower resolution. The first one can be extended to the reconstruction of multiple observations of a scenario, and the objective focuses on how to use all the complementary information for reconstruction. Most of them are spatial domain methods, such as projection onto convex sets (POCS) [12], the regularized methods [13, 14] or non-uniform interpolation [15].

SR methods based on single frame attempt to establish relationships between low and high resolution images through a prior learning process. One of the first works coincided with the rise of models based on neural networks in the 80s [16]. These algorithms learned the relationship between low quality images (usually a specific type of images: faces, fingerprints, etc.) and their high-resolution counterparts. With this learned knowledge they were able to improve the process of superresolution. In recent years, neuronal models are being used for the processing and recognition of images through the use of Convolutional Neural Networks (CNN). Although CNNs have been known for a few decades, deep CNNs are growing in popularity due to the good results obtained in classification problems with images [17]. This upswing in deep CNNs is mainly due to the appearance of new implementations with high computing capabilities using GPUs and the accessibility of the scientific community to the results of already trained models, for instance [18]. Deep CNNs have been used in SR problems [19, 20]. These neural models are able to learn the relationships between a low-resolution image with its high-resolution counterpart.

In recent years, the use of UAVs is growing in different areas of application. This growth is due to its versatility, flexibility, low-cost and minimized operational risk. However, UAVs have important limitations in terms of their autonomy. In applications where it is necessary to recognize a relatively large geographical area, it is possible to increase the height of the flight to maximize the observable area. This increase in height of the flight causes a loss of resolution of the image, so it seems logical to apply SR methods. It would be possible to use the methods of [19, 20] to improve the resolution of the UAV images, however, these models require high computing resources to be able to reconstruct the images in moderate times. These models would have to be executed in the embedded architecture of an UAV, which have limitations in their computing capabilities. We would encounter the problem of not being able to be executed on the on-board computer of the UAV, or the execution of the method could reach up to several minutes, so this solution would be infeasible. The aim of this paper is the proposal of a light model based on a deep CNN that allows to solve SR problems having as support the infrastructure embedded in an UAV.

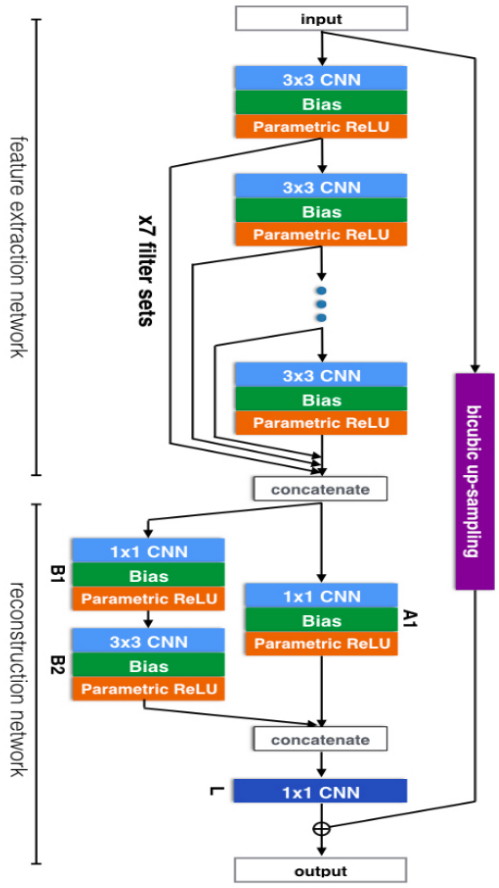
II. Deep Learning for Super-resolution images

In the specialized literature we can find works that use deep CNN applied to the problem of superresolution. The objective of these is to train a deep CNN capable of establishing the relationships between a low-resolution image with its high-resolution counterpart. Dong et al. [20] demonstrate how the deep CNN model can be viewed as a traditional sparse-coding-based SR method. The authors present a light model that obtains good restoration results in less time than other traditional methods. This model was called super-resolution convolutional neural network (SRCNN). Shi et al. [21] describe a deep CNN, which replaces the deconvolution layer of the SRCNN model with a proposal based on an efficient sub-pixel convolution layer which learns an array of upscaling filters to upscale the final LR feature maps into the HR output, in order to improve the performance.

The model chosen in this paper is the Deep CNN with Residual Net, Skip Connection and Network in Network (DCSCN) proposed by Jin Yamanaka, Shigesumi Kuwashima y Takio Kurita [22]. In comparison with other models, this model of SR achieves a fast and efficient computation, which is a fundamental requirement for a correct implementation in a UAV. The model is implemented in Tensorflow, so this will be the environment used in this work.

The DCSCN model works with the YCrCb image format. This format is a color space where each color is defined in terms of a luminance component and two chrominance components [23]. The Y channel represents the luminance, and the channels Cb and Cr represent the chrominance. Those last channels represent the blue and red colors respectively. All of these channels are defined between 0 and 255. The model does the super resolution to the entire image with a bicubic interpolation. The Y channel is the input of the neural network. The architecture of this model consists of a first series of convolutional layers grouped in cascade. With this type of grouping, the model achieved a better extraction of the characteristics of the image. Then, the result is taken and passed through convolutional networks grouped in parallel. This kind of aggrupation allows the model to reconstruct the image. The output of the model works as a mask that is added to the result of the bicubic interpolation. According to the authors, this architecture is optimized, since it achieves SR with a reduced number of layers, between 7 and 11, while other architectures requires more than 30 layers. In Figure 1 the DCSCN architecture is shown.

Figure 1: Architecture of the model described [24]



III. MODEL ADAPTED TO EMBED IT IN AN UAV

The model can be configured with some hyperparameters, like the number of layers of each section or the number of filters of each layer. The greater the hyperparameters, the greater the complexity of the model. Greater complexity does not always mean better results, but it does imply a longer execution time. To implement the model on a UAV, a short execution time will be required, so a model with little complexity should be used.

The authors propose hyper parameters that allow rapid execution without losing quality in the results. The hyperparameters used in this work are the following:

Table 1: Hyperparameters of the compact model

Feature extraction layers	Reconstruction images layers	Number of filters of feature extraction layers	Filters decay gamma	Number of filters of the reconstruction image layers
7	2	From 32 to 8	1,2	24 and 8

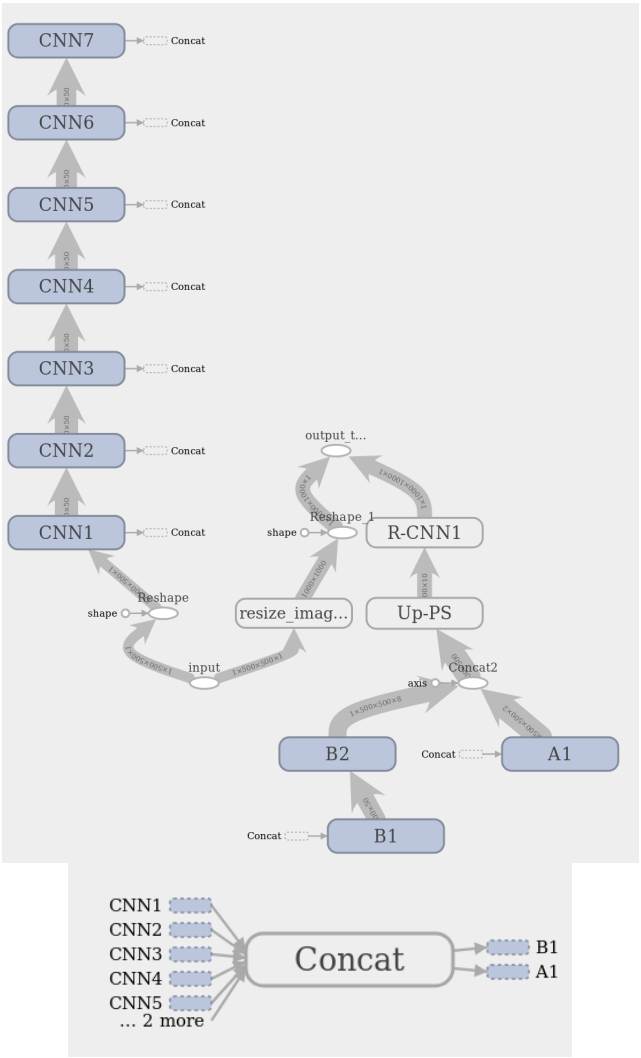
Once the model is suitable to execute on the desire device, the following step is to optimize it. The way to optimize a model is to remove all the operations that are not going to be used during an inference. For example, all of the operations related to the train phase, like the gradients and the optimizers. Those operations are not needed when doing an inference, so it will be removed from the graph.

The objective device of this paper is an UAV. It is assumed that this device has a Qualcomm chip, as these chips are one of the most popular in UAVs. These chips support artificial intelligence models, and an SDK called Snapdragon Neural Processing Engine (SNPE) [25] is provided to speed up the model execution processes on these chips. Qualcomm chips only support the .DLC format. The SNPE provides tools for the conversion of the models to this format, as well as other tools to do a benchmark. One aspect to keep in mind is that the SNPE converter does not support certain layers. In this case, the model contains some layers that are not supported, such as convolutional layer without bias. The solution to this problem is to replace these layers with other layers that are supported, and retrain the model. This may cause the results to get slightly worse

In Figure 2, the graph of the model adapted to be converted to the .DLC format is shown. As you can see, certain new layers have appeared in the model, such as those corresponding to the bicubic resolution that is performed on the Y channel. In addition, certain layers have been modified to be able to be supported in the required format.

With the converted model to DLC, it is time to develop an application to run the model. The SNPE provides functions in both C ++ and JAVA to load the model and execute it. The best way to test the correct implementation of the model for the Qualcomm chip is using a mobile phone with these components. For this paper, a BQ Aquaris X Pro 2 is used to do the benchmark.

Figure 2: Model changed to support .DLC format



IV. RESULTS AND DISCUSSION

A. *Experimental setup*

Once the model is obtained in .DLC and with an application that allows its execution, it is time to analyze the results. The chosen images represent images obtained by UAV to analyze the quality of the photos taken on those devices. These images represent those scenarios where UAVs are being used the most, since we want to demonstrate the improvement of quality in images using the proposed technique.

The other technique with which it is going to be compared is the bicubic interpolation, one of the most popular techniques at present to make superresolution. The scale factor chosen is x2, this means that the height and the width will be the double of the original image.

The set of images chosen aims to show different real scenarios where UAVs are needed to take images. The first image represents a sporting event, more specifically a soccer match, since in sport events UAVs are usually used to take images (Figure 3). The second image chosen is from a group of people, since the UAV can have the function of identifying people and the quality of the image is fundamental in this aspect (Figure 4). Finally, an image of a street has been used, where UAVs can also serve to monitor or identify certain things (Figure 5).

Figure 3: Image 1: Aerial view of a soccer match. [26]

Figure 4: Image 2: Aerial view of a group of people. [27]

Figure 5: Image 3: Aerial view of a street. [28]

The most popular metric used to measure SR results is the Peak Noise Signal Ratio (PSNR) [29]. This metric is used when evaluating the reconstruction of images, since it compares how similar an image is to another by means of the mean square error and is expressed in decibel (db). The formula of the PSNR is the following:

$$PSNR = 10 * log_{10}\left(\frac{R^2}{MSE}\right)$$

- R is the maximum value that pixels can take in the image format used, which will usually be 1 or 255.
- MSE is Mean Square Error, and is the mean square error between the two images to be compared.

The other metric that it is used is the Structural Similarity (SSIM). This metric measures the similarity between two images and is also used for evaluating the SR. More details about this metric can be found at [30]. The result is a value between -1 and 1, where 1 means that both images are identical.

The limitation of these metrics is that the resulting image is compared with the original image, so to obtain a better analysis of PSNR and SSIM, the original image should have a good quality.

B. Experimental setup

SR model (PSNR: **37.17db**, SSIM: **0.93**)

Bicubic (PSNR: 35.18db, SSIM: 0.91)

SR model (PSNR: **31.24db**, SSIM: **0.80**)

Bicubic (PSNR: 29.60db, SSIM: 0.75)

SR Model (PSNR: **41.57db**, SSIM: **0.99**)

Bicubic (PSNR: 36.64db, SSIM: 0.97)

In the following table, the resume of the results are showed:

Technique	PSNR Mean (db)	PSNR Stdev (db)	SSIM Mean	SSIM Stdev
SR model	**36,66**	5,18	**0,91**	0,1
Bicubic	33,81	3,72	0,88	0,11

As can be seen in the PSNR and SSIM values obtained, the proposed model manages to perform a SR without losing as much quality as occurs in the bicubic interpolation. Although the variability in the PSNR obtained is slightly

higher in the model, in all the images a higher value has been obtained than that obtained by the bicubic interpolation.

Regarding the execution time of this algorithm in the used device, the average execution time of the previous images was **334 ms**, with 121 ms of initialization time of the net. The execution of this net has been done at CPU due to the limitations of the selected device. The execution of GPU usually involves a slight increase in the initialization time, but the execution time will be significantly reduced.

V. CONCLUSIONS

As a final conclusion, an artificial intelligence model has been proposed with a better performance than the most popular techniques, such as bicubic interpolation. In addition, UAVs require that the execution of the SR be done in the shortest time possible, so this requirement has been met thanks to the implementation of the model in .DLC format using the SNPE offered by Qualcomm. Many other neural networks are prepared to run on SNPE, so having an algorithm ready to be executed with this tool in a short time reduces the total execution time on the UAV. In the neural networks such as object detection, the image quality is quite important for a better performance, so having a network capable of performing the SR in a short time and with good results is so important.

ACKNOWLEDGMENT

This work was funded by private research project of Company BQ and public research projects of Spanish Ministry of Economy and Competitivity (MINECO), references TEC2017-88048-C2-2-R, RTC-2016-5595-2, RTC-2016-5191-8 and RTC-2016-5059-8.

REFERENCES

[1] Allen Ferrick, Jesse Fish, Edward Venator and Gregory S. Lee "UAV Obstacle Avoidance Using Image Processing Techniques" Technologies for Practical Robot Applications (TePRA), 2012 IEEE International Conference on 23-24 April 2012

[2] G. Fasano, D. Accado, A. Moccia y D. Moroney, "Sense and avoid for unmanned aircraft systems" IEEE Aerospace and Electronic Systems Magazine, vol. 31, nº 11, pp. 82-110, 2016.

[3] Gerard Rankin, Andrew Tirkel, Anatolii Leukhin. "Millimeter Wave Array for UAV Imaging" MIMO Radar Radar Symposium (IRS), 2015 16th International 24-26 June 2015

[4] Farrel J.A., "Aided Navigation: GPS with High Rate Sensors", McGraw-Hill, New York, 2008.

[5] J. García, J.M. Molina "Analysis of Sensor Fusion Solutions for UAVs", Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA). 2018. Granada, España, 23-26 de octubre, 2018

[6] Gibbins, D., Roberts, P., & Swierkowski, L. (2004, December). A video geo-location and image enhancement tool for small unmanned air vehicles (UAVs). In Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004 (pp. 469-473). IEEE

[7] Haris, M., Watanabe, T., Fan, L., Widyanto, M. R., & Nobuhara, H. (2017). Superresolution for UAV images via adaptive multiple sparse representation and its application to 3-D reconstruction. IEEE Transactions on Geoscience and Remote Sensing, 55(7), 4047-4058

[8] Lei, J., Zhang, S., Luo, L., Xiao, J., & Wang, H. (2018). Super-resolution enhancement of UAV images based on fractional calculus and POCS. Geo-spatial Information Science, 21(1), 56-66.

[9] Greenspan, H. (2008). Super-resolution in medical imaging. The Computer Journal, 52(1), 43-63

[10] Mareboyana, M., & Le Moigne, J. (2018, April). Super-resolution of remote sensing images using edge-directed radial basis functions. In Signal Processing, Sensor/Information Fusion, and Target Recognition XXVII (Vol. 10646, p. 1064610). International Society for Optics and Photonics.

[11] Seibel, H., Goldenstein, S., & Rocha, A. (2017). Eyes on the Target: Super-Resolution and License-Plate Recognition in Low-Quality Surveillance Videos. IEEE access, 5, 20020-20035.

[12] Gao, J. J., Chen, X. H., Li, J. Y., Liu, G. C. & Ma, J. Irregular seismic data reconstruction based on exponential threshold model of POCS method. Appl Geophys. 7, 229–238 (2010).

[13] Liu, C., & Sun, D. (2014). On Bayesian adaptive video super resolution. IEEE transactions on pattern analysis and machine intelligence, 36(2), 346-360.

[14] Ng, M. K., Shen, H., Lam, E. Y., & Zhang, L. (2007). A total variation regularization based super-resolution reconstruction algorithm for digital video. EURASIP Journal on Advances in Signal Processing, 2007(1), 074585.

[15] Nguyen, N., Milanfar, P., & Golub, G. (2001). A computationally efficient superresolution image reconstruction algorithm. IEEE transactions on image processing, 10(4), 573-583.

[16] Mjolsness, E. (1985). Fingerprint hallucination. Diss. California Institute of Technology.

[17] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[18] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 248-255). Ieee.

[19] Kim, J., Kwon Lee, J., & Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).

[20] Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.

[21] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1874-1883).

[22] Yamanaka, J., Kuwashima, S., & Kurita, T. (2017, November). Fast and accurate image super resolution by deep CNN with skip connection and network in network. In Neural Information Processing (pp. 217-225). Springer, Cham.

[23] A. Weitzenfeld, "Espacio de color YCbCr", Cannes.itam.mx, 2018. [Online]. Available: http://cannes.itam.mx/Alfredo/Espaniol/Cursos/Robotica/Material/VisionAIBO.pdf

[24] Yamanaka, J., Kuwashim, S. and Kurita, T. (2018). *Model structure*. [image] Available at: https://arxiv.org/abs/1707.05425

[25] [2]"Snapdragon Neural Processing Engine SDK: Features Overview", Developer.qualcomm.com, 2018. [Online]. Available: https://developer.qualcomm.com/docs/snpe/overview.html.

[26] FramePool (2018). *Partido de Fútbol / Italia / Vista Aérea*. [image] Available at: http://footage.framepool.com/es/bin/113026,partido+de+f%C3%BAtbol,italia,vista+a%C3%A9rea/

[27] Marwood, S. (2015). *women-walking*. [image] Available at: https://www.womensaid.org.uk/information-support/what-is-domestic-abuse/women-walking/.

[28] Getty Images (2018). *TL, MS, HA Crowds and traffic on Hachiko Crossing, Shibuya / Tokyo*.

[29] Ni.com. (2013). *Peak Signal-to-Noise Ratio as an Image Quality Metric - National Instruments*. [online] Available at: http://www.ni.com/white-paper/13306/en/.

[30] Wang, Z., Simoncelli, E., Sheikh, H. and Bovik, A. (2004). [online] Cns.nyu.edu. Available at: http://www.cns.nyu.edu/pub/eero/wang03-reprint.pdf.