Counting Cattle in Argentina with AI and Satellite Images

Sandra Torrusio [1], Juan Gonzalez [2,3], Martin Errázquin [3], Abril Cereceda [1], M. Julia Cassano [1], Natalia Morales [1], Ricardo Saucedo [2]

[1] Faculty of Natural Sciences and Museum, National University of La Plata (UNLP), Av.122 & 60 street, La Plata (1900), Buenos Aires, Argentina – (storrusio, mariajulia.cassano, moralesnatalia)@fcnym.unlp.edu.ar, acereceda@gsuite.fcnym.unlp.edu.ar

[2] National Service of Agri-Food Health and Quality (SENASA), Av. Paseo Colon 367, Ciudad de Buenos Aires (1063), Argentina – jdgonzalez@senasa.gob.ar, rsaucedo@senasa.gob.ar

[3] Faculty of Engineering, National University of Buenos Aires (UBA), Av. Paseo Colon 870, Ciudad de Buenos Aires (1063), Argentina – merrazquin@fi.uba.ar

Keywords: AI, Counting Cattle, Natural Resources, VHR Satellite Images.

Abstract

From a specific need of the national institution responsible for plant and animal health in Argentina, SENASA, this collaborative work between the public sectors together with academia arises. The objective was the development of a Deep Learning algorithm for the detection of livestock, cows in particular, in very high-resolution satellite images (provided by the Argentine Space Agency (CONAE)) and its subsequent counting. The basic elements involved in the development of artificial intelligence are detailed, such as the selection and acquisition of satellite images, their very thorough preprocessing and labelling, details of the training stage and some forms of error quantification. The image database is made up of about 320 scenes (based on very high resolution (VHR) satellite data from the Pleiades and Pleiades NEO sensors, ©Airbus 2022, distributed by CONAE) of the Pampas and Patagonian regions in Argentina. Around 8,000 labels of different types of animals were generated, the most common were cows and sheep. Also labels that did not represent animals to contribute to training. Finally, some very promising preliminary results are presented, such as the average error in the count that was achieved was +/- four cows. The usefulness of these tools is reflected in better management of renewable natural resources linked to animal health issues, fiscal issues and within the framework of the 2030 Agenda and Sustainable Development Goals #12, #15 and #17.

1. Introduction

The operational monitoring of natural resources is a fundamental issue for public organizations in different levels such as national, provincial and local, as well as for productive sectors.

The study has as a main goal to develop a set of operative methodologies to count cattle with AI, already tested in other parts of the world, to satisfy a requirement of SENASA (National Service of Health and Agri-Food Quality), belonging to the Secretariat of BioEconomy of Argentina. The main function of SENASA is to ensure for animal and plant health and food safety, tasks that are often complex in a climate change scenario that we are going through, considering the 2030 Agenda and SDGs (Sustainable Developing Goals) #12 (Responsible Consumption and Production), #15 (Life in land) and #17(Parternships for the Goals). In some areas of the country, it is not easy to systematically identify livestock due to their wide extensions, presence of forests and plantations, hills and other landscapes. Automating the counting and detection of cattle is the main objective of the work. This record is necessary for health and fiscal controls of the cattle.

Detecting cattle using remote sensing technologies has been widely studied. Various types of sensors such as very high-resolution (VHR) satellite imagery, aerial photographs, and Unmanned Aerial Vehicles (UAV) are considered (Mücher et al., 2022). As the focus is on the well-being of cattle, one of the main interests lies in recognizing the posture of individual animals. The study found that VHR was not useful for their specific purposes. To support ecological analysis, other authors (Robinson et al., 2021) utilized Very High-Resolution (VHR) panchromatic imagery with a resolution of 0.3m/pixel to initially detect and subsequently count cattle and elk. Their dataset comprises 10,529 labelled points across 11 scenes from Maxar's Satellite catalogue, covering Point Reyes National Seashore, California, USA. The researchers applied various deep learning algorithms, which were evaluated using different performance measures specifically designed for the detection and counting of cows.

On the other hand, by using VHR imagery it focused on counting cattle in illegal cattle ranching located in Amazonas, Brazil, to estimate the impact of such animals on greenhouse gas emissions (Laradji et al., 2020). They analysed about twelve thousand patches of 250m x 250m, of which only 900 contained cattle. The common feature among the afore mentioned works is that all of them use deep learning algorithms for the vision branch.

In this work, we use a deep neural network to count cattle in different regions of Argentina, especially in the Pampas and Patagonian areas. Our main motivation is to provide a tool for assessing health risks in some isolated regions of the country.

Figure 1: Study area in Argentina. In pink Patagonian Region (South), in yellow Pampas Region (Center).

2. Materials and Methods

The qualitative specification of the counting problem can be summarized as: Given a high-resolution satellite image, estimate the order of magnitude of the number of cows present in the image.

There are different types of approaches, some are based on probabilistic estimates of pixel density, others are specifically designed filters that come from traditional digital image analysis. Some models work in a way that cannot be interpreted, generating an algorithmic output that does not allow subsequent explainability. In our case our approach is inspired by human logic, which is summarized in two steps:
1. Detect all instances of cows present in the image.
2. Count the number of detections.

Based on this approach, previously used successfully, a regressor based on object detection is proposed.
Unlike approaches that rely on point annotations (Laradji et al., 2018), which then present the problem of separating instances, we build a dataset with annotations by boxes (rectangles), which also make sense in the context of satellite images where it is possible to identify animals.

It is important to note that to count correctly it is crucial to detect correctly, that is, a low counting error requires a low detection error, at least in terms of classification (some level of error in box regression is acceptable).

For the object detection, in this particular case, the functionality is the following:
-input: a colour image in PNG format
-output: variable number of tuples (box, class, score), where box represents the four coordinates that make up the rectangle, class represents the class to which the content of the rectangle is associated, and score is a number between zero and one, which can be interpreted as a level of probability that the associated box can be associated with the required class.

The images are usually in RGB combination, partly inherited from the network backbone and which is standard. Each detection consists of an entity belonging to the provided image, determined by the box that contains it (typically the coordinates of two opposite vertices of the rectangle), the class to which it belongs (exactly one of the predetermined k possible classes) and the confidence you have in the prediction (a number from 0 to 1).

The satellite images collection constitutes the basis of the machine learning procedures. There the elements of interest are pointed out by human observers and the purpose of the algorithms is to imitate the behaviour present in the pointing, implicitly learning the characteristics of the objects of interest. In this case the image bank constitutes a set of satellite images potentially containing livestock, where it is labelled so that cows, horses, sheep, pigs and also guanacos present in the image are delimited with a rectangle on the image. Commonly to the people who perform this task are called annotators, they are expert in remote sensing, with good knowledge and management of tools like GIS, geospatial information and specific geo processes with the data. Each label is placed according to specific criteria that weigh different characteristics, taking into account the geographical area, the predominant type of livestock management, the surrounding vegetation (native or implanted), the relief, human facilities, among other features. The satellite images used in this work were Pleiades and Pleiades NEO (©Airbus 2022), (Figure 2 shows the provincial distribution of the scenes) in real colour combination (RGB), with a pixel size (spatial resolution) of 0.5m and 0.3m, respectively, and multispectral resolution (bands in visible and near infrared ranges), provided by CONAE in the framework of an agreement for non-commercial purpose.

This procedure is illustrated in Figures 3 and 4, where the original image is observed and an illustration of the annotation process. In technical terms each rectangle is encoded as a set of four coordinates that indicate the pixels corresponding to the corners of the box. In this work the annotation procedure was done using the free open-source web tool, Make Sense (Skalski, 2019).

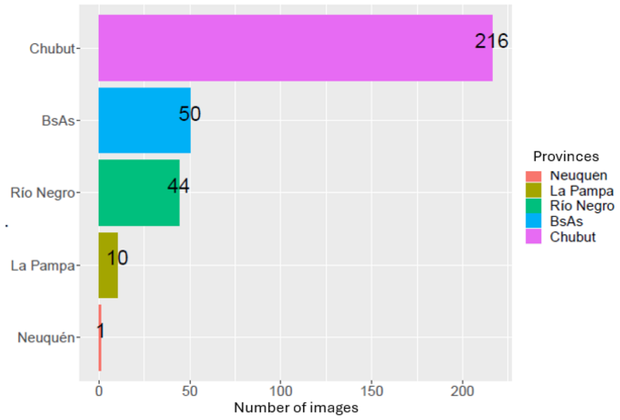

Figure 2. Distribution of number of scenes per provinces. Patagonian Region: Neuquen, Rio Negro, Chubut. Pampas Region: La Pampa, Buenos Aires (BsAs).

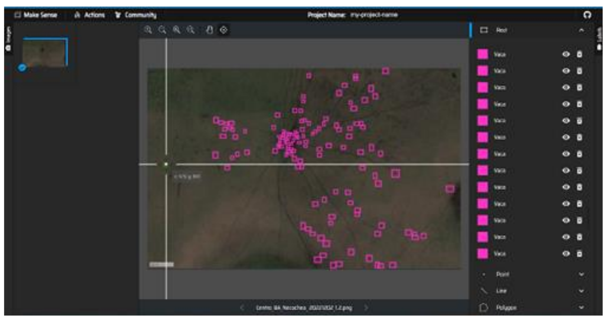

Figure 3. Visualization of labels in Make Sense on satellite image with coordinates x/y in each.

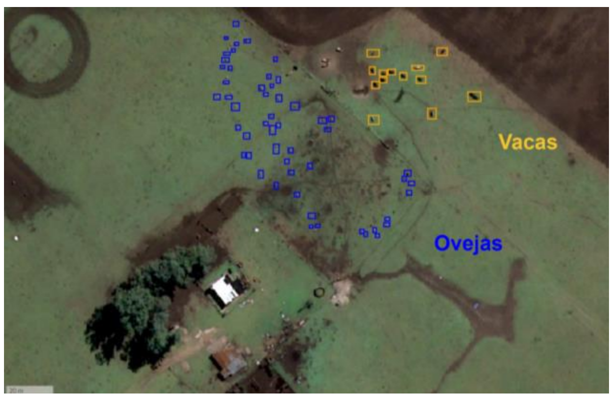

Figure 4. Manual Identification of cows (yellow) and sheep (blue) on the satellite image of VHR with the bounding boxes.

The model used is a deep convolutional neural network, it has two constituent parts which can be thought of as having two stages, in the first part, a so-called Region Proposal Network is formed. Its purpose is to locate the regions of interest of the image, in the later stage, each of those regions are combined in a way to define the entity label. In particular, we use the Faster R-CNN network (Ren et al., 2016), which is the result of a set of techniques that allow the acceleration of their predecessors R C-NN (Girshick et al., 2014) and Fast R-CNN (Girshick, 2015).

Since it has a very large number of parameters, the images necessary to train them from scratch should at least be of the order of the number of parameters. Instead, an estimation of the parameters is performed on a general-purpose image bank as a starting point, then training takes place based on slight modifications of this set of parameters where it is about adapting the network to the specific image bank. This type of approach is known as transfer learning.

For a counting method like the one presented, there are different ways of evaluating it, in particular, we take two widely used ones, due to their simplicity and interpretation.

Given an image, let N be the number of cows (ground-truth) and M be the number of detections made classified as cows. Note that N, M are positive integers or zero. They are defined as follows:

$$\text{MAE} = |N - M| \qquad \text{MAPE} = \text{MAE}/\max(N,1) \quad (1)$$

As a summary, MAE is the absolute error, whereas MAPE is the absolute error relative to the real number.

A dataset made up of 320 scenes in RGB format from various regions was used. Images were included without annotations, with cows and sheep annotated by experts; the test-set is formed by 70 images that were separated randomly for performance estimation and the remaining 250 were kept for training of the model (i.e the training set). The model training consisted of a fine-tuning procedure on a pre-trained Faster R-CNN model, a technique widely used in different works for detection of terrestrial fauna (Duporge et al., 2021). The selected implementation uses the Pytorch (Paszke et al., 2019) and Torchvision libraries.

The network is pretrained on the general-purpose image bank COCO dataset, (Lin et al., 2014), which consists of 330,000 images and its size is 25GB, (Elharrouss et al., 2022). Employing a pre-trained model involves accessing the network result in its final stage, without consumption of computational resources. However, the task of adapting the network to the specifically designed image bank is carried out in this work, with a very reduced consumption of resources compared to a case started from scratch. The model training was performed for a maximum of 60 epochs, the image bank is artificially increased, using a technique called data augmentation, which consists of generating new instances, for example by rotating or mirroring the images and their bounding boxes, respectively. In particular, horizontal reflections were used in a random manner (Random Horizontal Flipping with probability 0.5). In the training stage, the results of the algorithm in each iteration (epoch) were sensed with respect to the previously defined measures (MAE and MAPE).

The network output is set to detect a number of bounding boxes, each one has a score associated with the most probable class, where the class can be horse, cow, sheep, etc. If we focus on the cattle counting problem, we can take the boxes where the predicted class is cow. For a given box, a probability close to zero will mean little confidence that the determination is correct, while a probability close to one will mean greater certainty. The precise determination of which threshold to apply is considered taking into account the performance in the held-out data. To do so, in this work the values 0.3, 0.5, 0.7, 0.9 were explored.

In order to improve the results a post-processing was carried out. For cases in which two or more overlapping boxes correspond to a single cow, an ad-hoc post-processing technique called container suppression was designed. While for boxes of an area considerably larger than what a cow could occupy, a technique called feasible area was developed.

Container suppression consists in deleting any box which fully contains a smaller one; such procedure can be vectorized and only requires elementary arithmetic with bounding box coordinates. Feasible area consists of carrying out a prior statistical study to determine a cut-off point in which box size is decidedly implausible with respect to the present box area levels in the general sample.

Maximum Score is an additional post-processing technique that was applied after the two previously described ones, intended to reinforce the detection in the presence or absence of cows in the image. It consists of considering, for a given image, the maximum score of the detected box family. In cases where this amount is less than a threshold, the entire image is declared as having no cattle. Choosing the threshold value was based on held-out data, obtaining that the best performance was when the threshold was set at 0.8.

3. Results

We compare the results without and with post-processing over the test set. In the first iteration the error values are very high, and decreasing trends were observed as training progressed. The final model is achieved when the error reaches a relatively low value. This technique is known as early stopping and avoids a problem called overfitting present in many-parameter models (Goodfellow et al., 2016).

Figure 5 shows the behavior of the different error measures considered, in function of the iteration number (epoch). Previous iterations are orders of magnitude larger and not have been shown for a better interpretation of early stopping. Starting with iteration 10, is where the scale of measurements stabilizes. The evolution of performance measures shows a trend decreasing until a tentative convergence. The different initial thresholds considered are shown in curves of different colors. The choice of network weights is based on locating an iteration and a threshold that are considered sufficiently small compared to the rest. Based on the comments previous, with prioritization on the MAE error measure, epoch 12 is chosen, with the threshold 0.3. The approximate MAE of this configuration is 4.3. From now on this error measure will be reduced later of applying the post-processing mentioned.

With post-processing the results of the trained model are presented, which has the MAE as its governing error measure. While the MAE tries to give an idea of the average counting error of the number of cows correctly detected, the MAPE focuses on the percentage error. This metric is very easily inflated by 0-cow images, where each false positive increases the percentage error by 100%.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-459-2024 | © Author(s) 2024. CC BY 4.0 License.

In turn, the three post-processing techniques are indicated with suffixes af and cs, ms indicating container suppression, feasible area and maximum score respectively. Table 1.

Model	**Original**	**CS**	**CS+AF**	**CS+AF+MS**
MAE	4.30	4.03	4.00	3.77
MAPE	2.21	1.99	1.97	1.74

Table 1. Results of the initial model and performance measurements, after the post-processing carried out. (cs: Container Supression, af: Feasible Area, ms: Maximum Score).

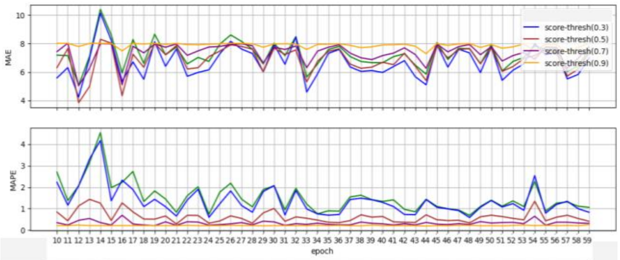

Figure 5. Evolution of the performance measures considered (MAE, up, and MAPE, down), throughout the iterations (Epochs) of the Deep-Learning algorithm. In different colors threshold curves corresponding to the minimum score used in determining whether individual boxes of the empty class are considered as such.

The diagonal comparison is presented as a way to visualize the behavior of the network in the test sample. The final result of the training is seen in Figure 6. Each point represents the result of the image, where on the abscissa axis (x axis) is the real number of cattle present, while the ordinate axis (y axis) shows the number predicted by the network, including its stage of post processing. In a completely correct determination, the points should be placed on the line identity (y = x), located at 45 in gray dotted lines. This line separates the coordinate plane into two clearly defined regions, a point far from said line means overestimation or underestimation according to whether they are located in the region of the upper or lower plane, respectively. Additionally, two orange lines defined by y = 10x and y = 0.1x are shown, the points that fall within this region are those where the determination of the model prediction fits, in order of magnitude the number of cows present. Subject to the condition that there are cows present in the image, all the points have remained in that region, which indicates that the model is capable of determining presence with correct orders of magnitude. On the other hand, it can be seen that there is a set of images whose determination is made in such a way that they indicate presence of livestock where the image does not contain cows, these points are represented on the ordinate axis and although their number has been reduced after the Maximum Score post-processing, the figure shows that it has not corrected this defect in its entirety. Finally, green dotted lines are given by y = 2x and y = 0.5x, the objective of incorporating them is to show the order of magnitude of the error, but with a refinement in the level of precision, in effect, it tells us that the points that are enclosed by these lines are estimated with a maximum error of double or half of the true count.

The following figures 7, 8, 9 and 10 show specific examples of the algorithm application in different scenarios according to the area, presence of animals and features of the images (textures, shapes, tones, relief, infrastructures, etc.).

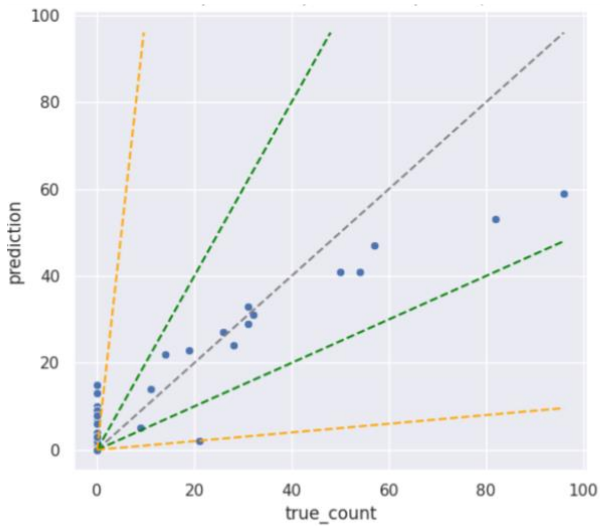

Figure 6. Final result of the training. Each point represents the result of the neural network for an image, where on the abscissa axis (x axis) is the actual number of cattle present, while on the axis of the ordinates (y axis) shows the number predicted by the network, including its post-processing stage. In a totally correct determination, the points must be placed on the identity line (y = x), located 45 in gray dotted lines. The dotted orange lines have equations y = 0.1x and y = 10x, while that the green dotted y = 0.5x and y = 2x, both were included to determine the error in the order magnitude of each image with different levels of refinement.

Figure 7. Example of high error when the there are no cows.

Figure 8. Example of low prediction error when the livestock presence is low. In orange (up) the box with manual annotation, and in red (down) the prediction of the trained neural network.

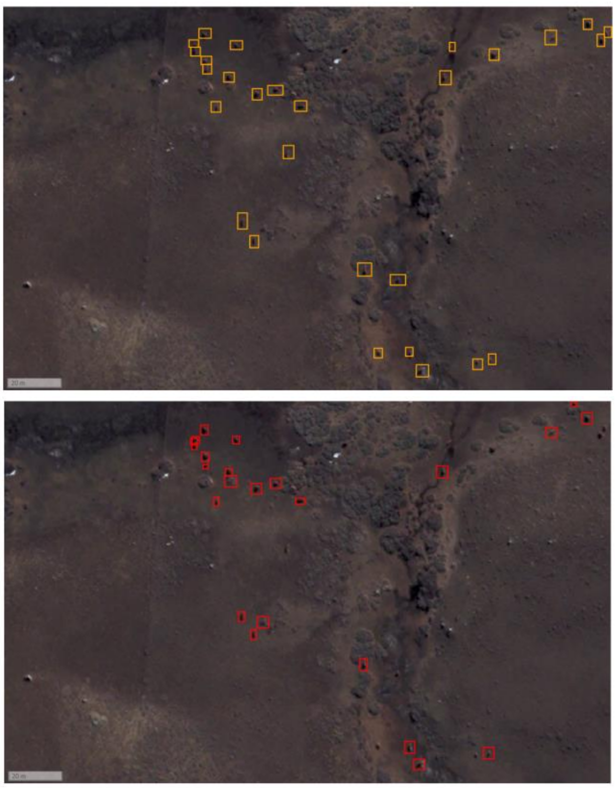

Figure 9. Example of low prediction error when the livestock presence is medium. In orange (up) the box with manual annotation, and in red (down) the prediction of the trained neural network.

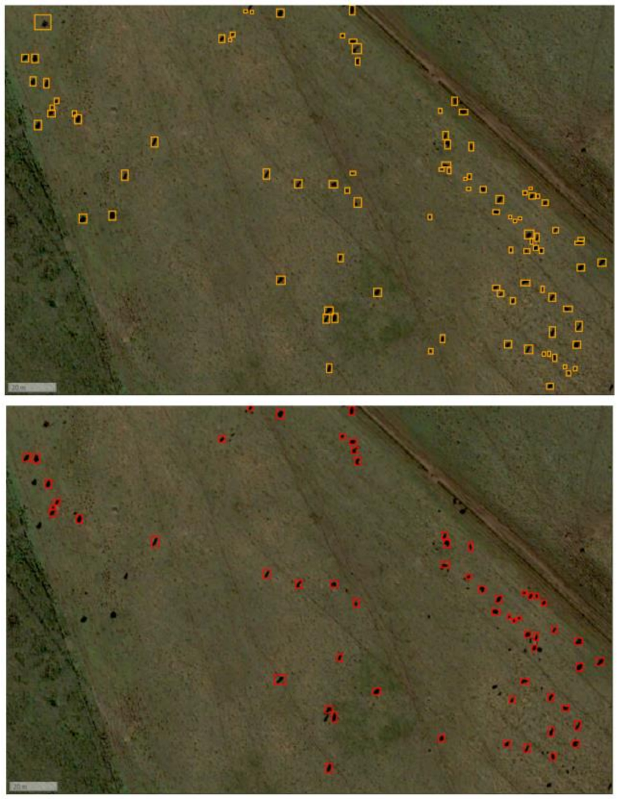

Figure 10. Example of high prediction error when the presence of livestock is high. In orange (up) the box with manual annotation, and in red (down) the prediction of the trained neural network.

4. Main Conclusions

Based on a relatively small set of data, it has been possible to estimate in order of magnitude the number of cows present in a satellite image. The error in absolute terms of the number of animals for the region of interest is dependent on the actual number of livestock. As an example, for an image containing few livestock it may be less than four cows (MAE ≈ 3.7).

The performance of the network can be improved based on two strategies, the first is by growing the image bank. The second one involves employing a set of problem-specific, post-processing techniques addressing the network shortcomings.

We would like to highlight the importance of the cooperation between public organisms and the scientific sector which made this work possible: SENASA (National Service of Agri-Food Health and Quality), CONAE (National Commission of Space Activities) in its role of provider of satellite images, the Faculty of Natural Science and Museum (UNLP) and Faculty of Engineering (UBA), each one with their experts in different disciplines have contributed to this paper, one of the first developed in the country in this topic (Ocholla et al., 2024).

If we now consider the environmental aspects of these developments, it is probably that we can use them for other kind of animals, wild and domestic, and promote the use of technology and AI methods to improve the management and habitats conservation for a sustainable developing.

Within the framework of the United Nations (UN) Agenda 2030, the SDS (Sustainable Development Goals) and the targets that each country proposed to promote the transformation that the planet needs in relation to climate change, we have to take into account that today technology focused on concrete solutions is essential. We have mentioned that this work can contribute mainly to three of the 17 SDGs: #12, #15 and #17. The #12 *“Guarantee sustainable consumption and production patterns”*: if we consider that livestock production is one of the main emitters of GHG (Green House Gasses), that many times this productive activity advances on natural habitats such as grasslands, native forests, wetlands, which fulfill specific ecosystem functions, we see that tools like this development can help to decision-makers in developing countries, countries with economies in transition or with low-income to increasingly move towards sustainable and innovative development and preserve food security to transforming the current environmental and socioeconomic situation.

With respect to SDG #15 *“Life on Terrestrial Ecosystems”:* these advances in AI and remote sensing as is presented in this paper will contribute to better understanding the distribution of native fauna and possibly sensing it with field support and knowledge from experts of different disciplines.

Finally, SDG #17 *“Strengthen the means of implementation and revitalize the Global Alliance for Sustainable Development”:* in this framework it is important, as already mentioned, cooperation between institutions/countries/regions, recognizing that today the solutions and transformations need of multiple stakeholders to mobilize and share knowledge, technology, expertise and even financial resources to achieve the SDGs, particularly in developing countries.

5. Acknowledgements

The authors thank the support received by SENASA and its staff (Dirección de Tecnologías de la Información), by the National University of La Plata (UNLP) (Dr. Javier Diaz), by Faculty of Natural Sciences and Museum (Dr. Eduardo Kruse) and by the National University of Buenos Aires (UBA).

Another special thanks to CONAE (National Commission of Space Activities) (Nathalie Horlent, Aldana Bini) for providing satellite data that is not easy to acquire.

Thank you to the reviewers for their contributions.

6. References

Duporge, I., Isupova, O., Reece, S., Macdonald, D. W., Wang, T., 2021: Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. *Remote Sensing in Ecology and Conservation*, 7(3), 369–381. doi.org/10.1002/rse2.195.

Elharrouss, O., Akbari, Y., Almaadeed, N., Al-Maadeed, S., 2022. Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. doi.org/10.48550/arXiv.2206.08016.

Girshick, R., 2015. Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440-1448, doi.org/10.1109/ICCV.2015.169.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. doi.org/10.48550/arXiv.1311.2524.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT press. https://www.deeplearningbook.org/

Laradji, I. H., Rostamzadeh, N., Pinheiro, P. O., Vazquez, D., Schmidt, M., 2018. Where are the blobs: Counting by localization with point supervision. doi.org/10.48550/arXiv.1807.09856.

Laradji I. H., P. Rodriguez, F. Kalaitzis, D. Vazquez, R. Young, E. Davey, A. Lacoste, 2020. Counting cows: Tracking illegal cattle ranching from high-resolution satellite imagery. doi.org/10.48550/arXiv.2011.07369.

Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Doll’a r, P., Zitnick, C. L., 2014. Microsoft COCO: common objects in context. CoRR, abs/1405.0312.doi.org/10.48550/arXiv.1405.0312.

Mücher, C. A., Los, S., Franke, G. J., Kamphuis, C., 2022. Detection, identification and posture recognition of cattle with satellites, aerial photography and UAVs using deep learning techniques”. *International Journal of Remote Sensing*, vol. 43, no. 7, pp. 2377–2392. doi.org/10.1080/01431161.2022.2051634.

Ocholla, I. A., Pellikka, P., Karanja, F. N., Vuorinne, I., Odipo, V., Heiskanen, J., 2024. Livestock detection in African rangelands: Potential of high-resolution remote sensing data. *Remote Sensing Applications: Society and Environment*. doi.org/10.1016/j.rsase.2024.101139.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’AlchÃ© Buc, E. Fox, y R. Garnett (Eds.) *Advances in Neural Information Processing Systems 32*, (pp. 8024–8035). Curran Associates, Inc. doi.org/10.48550/arXiv.1912.01703.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. doi.org/10.48550/arXiv.1506.01497.

Robinson C., Ortiz, A., Hughey, L., Stabach, J.A., Ferres, J. L. M., 2021. Detecting cattle and elk in the wild from space. doi.org/10.48550/arXiv.2106.15448.

Skalski, P., 2019. Make Sense. https://github.com/SkalskiP/make-sense/