Artícle **Journal of Technological Development**
December 2022, Vol.6 No.18 12-20

Training dataset generation for semantic segmentation utilized unmanned aerial vehicles

Generación de datos de entrenamiento para segmentación semántica usando vehículos aéreos no tripulados

LÁRRAGA-ALTAMIRANO, Hugo René†*, HERNÁNDEZ-LÓPEZ, Dalia Rosario, PIEDAD-RUBIO, Ana María and AVILÉS-GUERRERO, Carlos Leonardo

Tecnológico Nacional de México, Campus Ciudad Valles, Mexico.

ID 1st Author: *Hugo René, Lárraga-Altamirano* / **ORC ID:** 0000-0001-8258-9418, **Researcher ID Thomson:** T-2296-2018, **arXiv Author ID:** Hugo_Larraga, **CVU CONACYT ID:** 626539

ID 1st Co-author: *Dalia Rosario, Hernández-López* / **ORC ID:** 0000-0002-2751-5886, **Researcher ID Thomson:** T-2470-2018, **arXiv Author ID:** DaliaHernandez, **CVU CONACYT ID:** 536472

ID 2nd Co-author: *Ana María, Piedad-Rubio* / **ORC ID:** 0000-0003-1258-0383, **Researcher ID Thomson:** T-2477-2018**, arXiv Author ID:** ampiedad, **CVU CONACYT ID:** 732279

ID 3rd Co-author: *Carlos Leonardo, Avilés-Guerrero* / **ORC ID:** 0000-0003-0316-3019, **Researcher ID Thomson:** GPS-9883-2022, **arXiv Author ID:** carlos#15, **CVU CONACYT ID:** 1244803

DOI: 10.35429/JTD.2022.18.6.12.20 Received July 20, 2022; Accepted December 30, 2022

Abstract

The present work focuses on the generation of training data for semantic segmentation using unmanned aerial vehicles, data acquisition was carried out, which were later processed, cut and labeled, generating the following results: datasets of sugar cane crops sugar for data analysis and field decision making. Use of geographic information system tools to increase the resolution of orthomosaics through interpolation. Design of a multivariate linear regression model to create a representative orthomosaic of the blue band.

Remote sensing, DataSet, UAV, Segmentation

Resumen

El presente trabajo expone una metodología para la construcción de datos de entrenamiento (dataset) en clasificadores supervisados utilizados para procesos de segmentación semántica. Los resultados muestran la adquisición de imágenes de cultivos de caña de azúcar obtenidas a través de vehículos aéreos no tripulados; la conformación de ortomosaicos, uso de sistemas de información geográfica para aumento de resolución por interpolación; el diseño de un modelo de regresión lineal multivariable para generación de datos sintéticos; el etiquetado de los mismos y el recorte de ortofotos.

Percepción remota, DataSet, VANT, Segmentación

Citation: LÁRRAGA-ALTAMIRANO, Hugo René, HERNÁNDEZ-LÓPEZ, Dalia Rosario, PIEDAD-RUBIO, Ana María and AVILÉS-GUERRERO, Carlos Leonardo. Training dataset generation for semantic segmentation utilized unmanned aerial vehicles. Journal of Technological Development. 2022. 6-18:12-20.

* Author's Correspondence (hugo.larraga@tecvalles.mx)
† Researcher contributing first author.

December 2022, Vol.6 No.18 12-20

Introduction

By means of remote sensing techniques it is possible to obtain information about an object or phenomenon without being in direct contact with it (Lira, 2003) (Lillesand & Kiefer, 2007). Remote sensing makes it possible to classify the types of crops that appear in the images and helps to know the phenological state of each crop through spectral signatures, so it can be used to: obtain crop surface statistics, evolution of the areas occupied by crops, monitoring the vegetative state of crops, management of irrigation water and control of its application, as well as precision agriculture (Arenas, 2016).

Remote sensing combines several elements for its application, such as: the platform (satellite, aircraft, or unmanned aerial vehicles known as UAVs or drones) (INEGI, 2014); the object to be observed (surface of the earth) and sensors (satellite system or sensor, -camera or video camera-) (Chuvieco, Teledetección ambiental: la observación desde el espacio, 2002).

In order to obtain quantifiable data from remotely sensed images, it is necessary to carry out processing that facilitates their conversion to biologically meaningful units and indices. This is achieved by processing and analysing the different variables represented by the pixels from different bands according to the sensors that have been used (Arivazhagan, 2013) (Bock, 2010).

For example, the infrared band of the electromagnetic spectrum has an important role in crop monitoring, as it provides information related to the biochemical processes of plants, so these values are used in the calculation of different vegetation indices such as the Normalised Difference Vegetation Index (NDVI) (Rueda, 2015). The analysis of these images has been useful to detect early stages of infections or diseases, nutrient deficiency and crop dehydration. This implies improvements in intervention, prevention and control of various problems associated with crop management. (Mahlein, 2016) (Grupta & Ibaraki, 2014).

Two techniques are used in land cover classification of remotely sensed images: supervised, the user preliminarily recognises known regions of interest in the land area, and the chosen algorithm extrapolates these spectral features to other regions of the image, thus performing the classification (Castillejo-Gonzalez, et al., 2009) and unsupervised classification aims to group cases by their relative spectral similarity, without field sampling (Foody, 2002).

Supervised classifiers have shown highly effective results in land cover classification, especially those in the area of deep learning (Kim, Lee, Han, Shin, & Im, 2018). Neural networks, particularly convolutional neural networks (CNNs) as deep learning techniques, are an effective tool for characterising, modelling and predicting a large number of non-linear processes with adequate results in decision making required in complex agricultural problems, including: prioritising and classifying products, pattern recognition, crop prediction and physical changes of their products (Figueredo-Ávila & Ballesteros-Ricaurte, 2016).

One of the challenges of supervised learning and consequently of CNN networks is that they require a large diversity of training data to achieve reliable accuracy levels when classifying an image (Hu, Luo, & Wei, 2020). In addition, data labelling is also a challenge of the training process of a supervised classifier. For this purpose, manual intervention, automatic or semi-automatic labelling or synthetic data emulating images are used (Donyavi & Asadi, 2020).

Therefore, this paper proposes a procedure for generating a dataset of images corresponding to sugarcane crops. It documents the acquisition of images using a UAV and a multispectral camera, the construction of RGB and multispectral orthomosaics, their adjustment to achieve a higher resolution, and the creation of a synthetic spectral band. Additionally, a binary labelling is proposed with the intention of identifying only the sugarcane crop on an orthomosaic. Finally, a cropping algorithm is described to extract sub-images from the orthophoto that integrate the dataset. It is important to mention that this data will be used as input for the training of a CNN to achieve a semantic segmentation.

The basic process of semantic segmentation is to pre-process the image to extract the spatial or spectral features that will allow distinguishing each pixel and predict the belonging to a specific class (Zabawa, et al., 2020).

Metodology

Supervised classifiers exposed in machine learning and deep learning techniques, used to distinguish vegetation types such as: Random Forest and Convolutional Neural Networks respectively, require a large amount of training data to perform the task of classifying efficiently (Di Cicco, Potena, Grisetti , & Pretto, 2017). Moreover, the input data should be suitable for the pattern recognition task that is intended to be designed, i.e., diversity, labelling, cleanliness of the data are desirable features in the dataset integration task (Donyavi & Asadi, 2020).

Data acquisition

The use of UAVs has had great impact on agricultural activities, given their ability to observe through specialised sensors details in crops, which the human eye could not with the naked eye (Tripicchio , Satler , Dabisias, Ruffald, & Avizzano, 2015) (Tripicchio , Satler , Dabisias, Ruffald, & Avizzano, 2015). Flight plans are designed considering several factors to ensure the successful acquisition of the images avoiding the risk of an accident, the following factors are considered:

- Weather conditions: sufficient light, avoid cloudy or rainy days.

- Flying hours: morning flights, avoiding excessive solar radiation, between 8 and 10 am.

- Photogrammetric parameters: flight height up to 100 m, vehicle speed 10 m/s, splicing percentage 80-85%, camera angle 90 degrees.

- Equipment calibration: both the UAV and the multispectral camera should operate with optimal battery and calibration levels (Lyu , Vosselman , Xia , & Yilm, 2020).

Processing of the acquired images

The sensor used is a Sequoia multispectral camera from the manufacturer Parrot, figure 1, equipped with an RGB lens and four bands: red, green, near infrared (NIR) and red limit. The RGB lens has a resolution of 16 Mpx and the rest of the bands 1.2 Mpx. In addition, the sensor performs a radiometric calibration automatically.

Figure 1 Sequoia multispectral camera
Source: https://geoinstrumentoscol.com/product/parrot-sequoia/

The images obtained from the sugar cane crops are processed with Pix4D Mapper software. RGB and Multispectral projects are generated by applying the Ag RGB and Ag Multispectral templates respectively. In such a way that one RGB orthomosaic and 4 orthomosaics are created for each of the red, green, NIR and boundary red bands (Olsson, et al., 2021).

The spatial resolution (GSD, ground sample distance) is an important element when it comes to analysing the content of an image. The height of flight and the type of sensor determine the sharpness with which details in the image are visualised (Lee, Son, & Kim, 2022). To improve the resolution of orthomosaics, i.e. decrease the GSD value, interpolation operations are performed that decrease the pixel size without losing information (Guarneri & Weih, 2010).

There are no ground control points (GCP) to support the rectification of orthoimages, therefore, coming from different lenses, the angles and position of the images are not equal, which causes the RGB orthomosaic and multispectral ones to be misaligned (Vassilopoulou, Hurni, & Dietrich, 2002).

Therefore, the blue band of the RGB orthomosaic cannot be used with the individual red and green bands. However, it is possible to generate a synthetic orthomosaic from the information contained in the 3-band orthomosaic. It is observed that the relationship between the red and green bands compared to the blue band is linear and considering that the images have been captured under the same atmospheric conditions, a multivariate linear regression model is designed to be used to create a new orthoimage with reflectance values in the blue band of the electromagnetic spectrum (Li, Hua, & Lu, 2021).

Orthomosaic cropping

The constructed orthoimages are arrays of thousands or millions of pixels that store digital reflectance values. Operating with images of these dimensions to train a classifier to perform semantic segmentation is not computationally possible. Therefore, orthomosaics are divided into 256 x 256 px subimages, generating a sufficient amount of training data for each cropped orthoimage (Ling, Cheng, Peng, Zhai, & Jiang, 2022).

Pixel tagging

This is one of the most challenging phases in dataset creation, especially when working on surface analysis and the training dataset is oriented towards supervised classifiers. The diversity of objects within an image and the spectral similarity between classes makes it a difficult task to label pixels (Lei & Peng, 2020).

In this study, a training dataset is integrated for binary classification. For this purpose, the NIR band is particularly valuable in the study of vegetation cover, such as crops in this case. High NIR values are related to the presence of vegetation, conversely the absence of vegetation is denoted by low NIR values (Gašparović, Zrinjski, Barković, & Radočaj, 2020).To label the two classes of interest a thresholding method is used, value one (1) corresponds to crop and value zero (0) to any other element which can be: barren soil, asphalt, buildings, among others (Gao, Xiao, & Jia, 2020).

Results

Data acquisition

A DJI Matrice 600 Pro drone was used to perform the flight, it was equipped with a Parrot Sequoia camera with light sensor and additional power source, figure 2.

Figure 2 Dji Matrice 600 Pro
Source: Own elaboration

The efficiency of the batteries is 22 min with a 100% charge considering 30% of the energy for the end of the flight. Both the drone and the camera were calibrated before the take-off operation, it was confirmed that the sensors, radio frequency and firmware were in adequate conditions, additionally the images of the calibration plate were captured for later correction, see figure 3.

Figure 3 Sequoia Radiometric Calibration Plate
Source: Own elaboration

The autonomous flight was programmed with the freely distributed Pix4D Capture software for Android and IOS platforms, figure 4. The configurable parameters are height, 80 m; vehicle speed 10 m/s; splice 80%. The multispectral camera was placed at a 90 degree angle to the ground with a capture time of 1.5s.

ISSN 2531-2960

Figure 4 Pix4D Capture, flight planning
Source: Own elaboration

Processing of acquired images

For the construction of the orthomosaics, Pix4D Mapper software was used to process two project types: RGB (figure 5) and multispectral.

Figure 5 RGB Orthomosaic
Source: Own elaboration

The latter results in orthophotos for each of the red, green, near infrared and boundary red bands (figure 6). The integration of each set of orthophotos per band into an orthomosaic was carried out with the help of the geographic information system (GIS) QGIS.

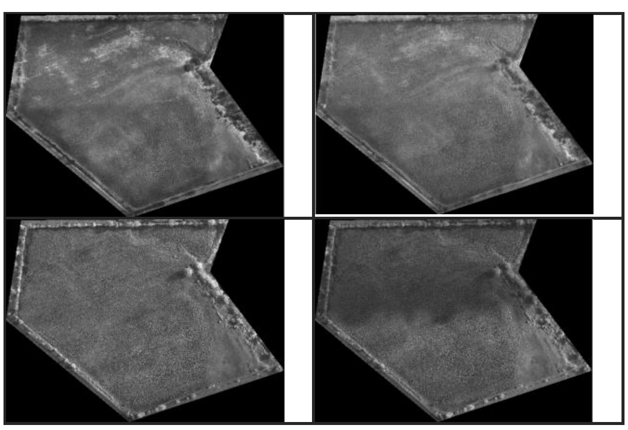

Figure 6 Orthomosaics red, green, near-infrared and boundary red bands
Source: Own elaboration

The GSD value of the orthomosaics differs by the resolution of the RGB lenses (16 Mpx) and the rest of the bands (1.2 Mp), being 2.22 and 8.81 respectively. QGIS through the georeferencing tool allows geometric corrections based on control points, so it can be used to decrease or increase the GSD of the orthophoto.

Several geographic points of the RGB orthomosaic were taken as control points to increase the resolution of the multispectrals, using a polynomial transformation, the nearest neighbour algorithm for resampling and the target resolution was set. After this processing the dimension and resolution of the orthomosaics is the same, width 26280 px and height 19223 px, with a GSD value of 2.22 cm/px, see figure 7.

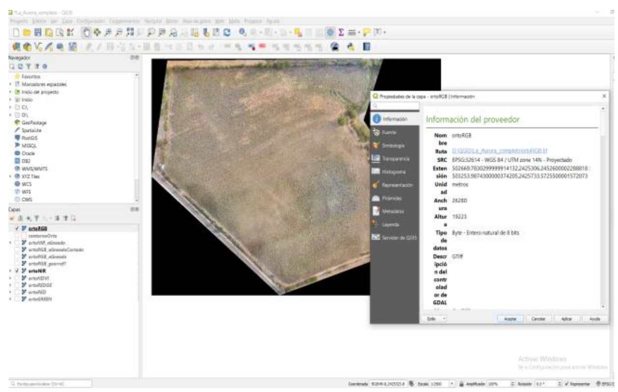

Figure 7 Orthomosaic processing in QGIS
Source: Own elaboration

Since the RTK equipment is not available to generate control points and thus obtain aligned RGB and multispectral orthomosaics, a synthetic blue-band orthomosaic was created to complement the red and green sensor of the Sequoia camera. A Multiple Linear Regression model was designed, given the linearity of the blue spectrum compared to the red and green, as shown in figure 8.

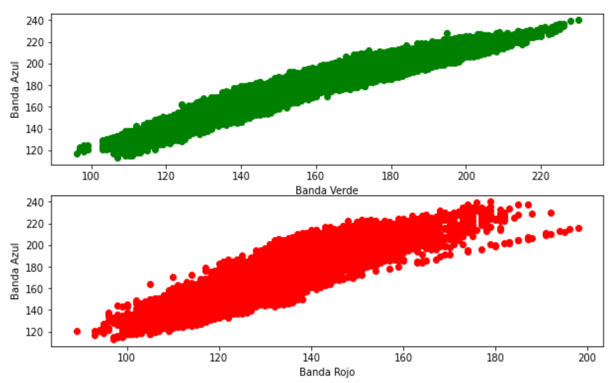

Figure 8 RGB orthomosaic scatter diagram
Source: Own elaboration

ISSN 2531-2960

The model was trained with the RGB orthomosaic data, as it is created under the same climatic and temporal conditions as the individual red and green orthomosaics of the Sequoia camera. The model recorded the following result parameters:

- Value of slopes or coefficients "a": [0.82965658 0.30154856]

- Value of intersection or coefficient "b": 2.984854941495797

- Value of the intersection or coefficient "b": 2.984854941495797

- Model accuracy: 0.974906969683250053

In this way, the model is applied by taking as input the red and green (individual) pixels and predicts the value corresponding to the blue pixel, creating an orthomosaic synthetically, see figure 9.

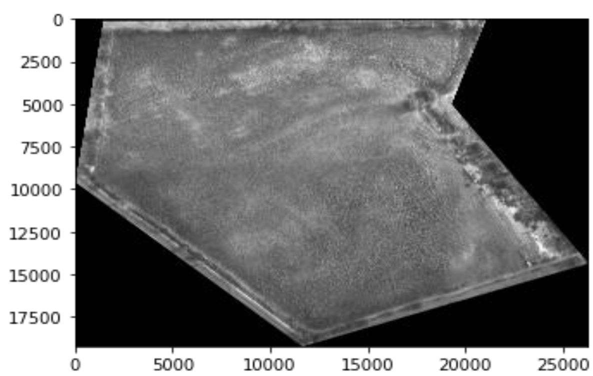

Figure 9 Blue band synthetic orthomosaic
Source: Own elaboration

Cutting out orthomosaics

Orthomosaics of sugar cane crops are extremely large images in the order of millions of pixels (26280 x 19223). Supervised learning in the training phase requires computationally processable inputs due to the number of operations performed during training. A program was designed in Python 3.8.8 using the Anaconda platform and the Jupyter NoteBook tool, whose objective is to extract sub-images with dimensions of 256 x 256 px that function as inputs to a supervised classifier. The algorithm works as a window that slides from the left margin of the orthoimage to the right boundary and from the top to the bottom, see figure 10.

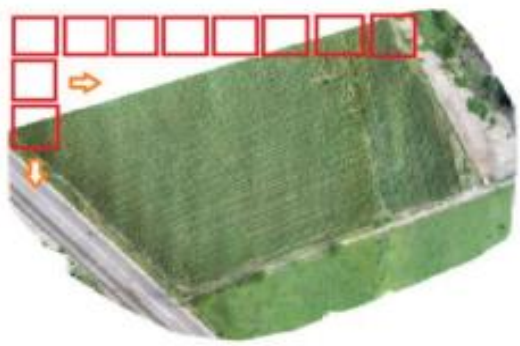

Figure 10 Operation of the trimming algorithm
Source: Own elaboration

Pixel tagging

To generate the binary images of the sugar cane crops that the classifier uses to adjust the learning model, the cropped NIR orthomosaic was taken. The segmentation method implemented is thresholding, i.e., it is required to determine a value that works as a point of comparison, those values less than zero and those greater than 1.

The Otsu method was applied to determine the threshold in each sub-image, so that 7650 threshold values were obtained. The frequency graph (histogram) shows that from value 80 onwards the highest frequencies appear, which belong to the vegetation, mostly to the sugar cane crop, as can be seen in figure 11.

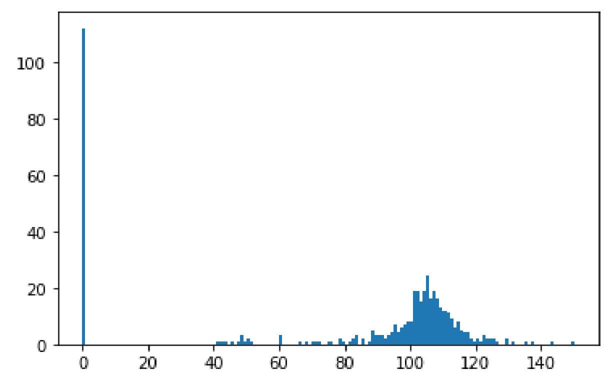

Figure 11 Histogram of thresholds NIR images
Source: Own elaboration

The result of binarisation with a threshold of 80 is effective, higher values (whites) represent vegetation while zero values identify other bodies in the image, see figure 12.

ISSN 2531-2960

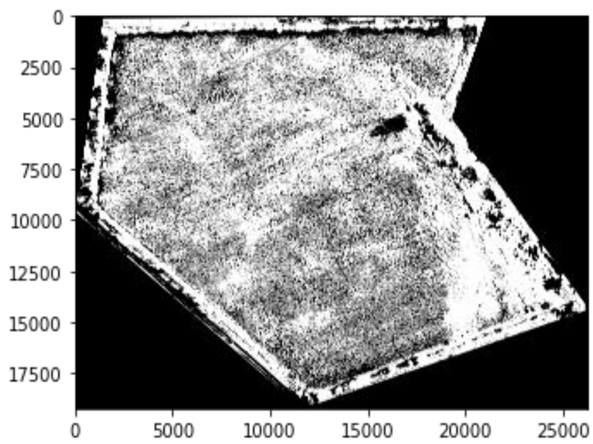

Figure 12 Binarised NIR orthomosaic
Source: Own elaboration

However, human intervention is necessary to edit the binary image. Although the NIR differentiates the vegetation from the rest of the image bodies, the image must show only the sugar cane. For this reason, pixels that do not represent the crop are removed using digital editing software, as shown in figure 13.

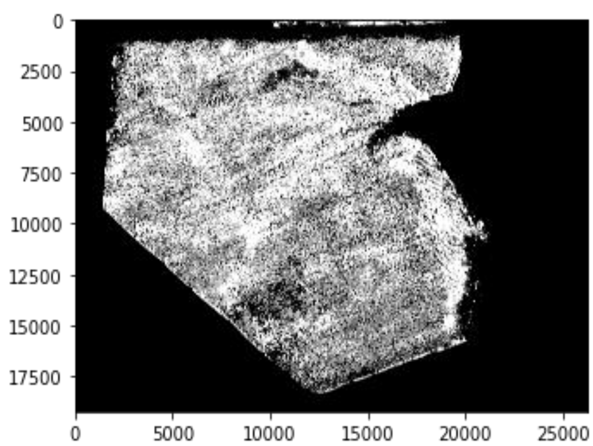

Figure 13 Binarised NIR orthomosaic, edited
Source: Own elaboration

Subsequently, it was processed with the clipping algorithm, like the rest of the orthomosaics. Images with binary labels are considered as part of the dataset and can be used to train convolutional neural networks dedicated to semantic segmentation.

Discussion of the results

It would be advisable to generate RTK control points in the photogrammetric flights to assist in the processing of the images and thus create rectified and aligned orthomosaics.

The multivariate linear regression model designed is suitable only for one flight, i.e. the model needs to be retrained with each RGB orthophoto to create the missing synthetic blue band, as atmospheric conditions change. The Otsu thresholding method does not effectively distinguish the other vegetation class and the crop, given their spectral similarity in the NIR band, other segmentation methods should be implemented to have better results.

Future work

Work is to be done on the multi-labelling of the images, considering 5 classes: null images, barren soil, asphalt, other vegetation and sugar cane cultivation.

The constructed dataset will be used for semantic segmentation tasks using convolutional neural networks. The objective with the binary training data is to identify the sugar cane crop in an orthomosaic. By having images with several labelled classes, the segmentation result would be a function of the number of classes.

Conclusions

The training dataset consists of 7650 images of sugar cane crops per lens of the Seqouia multispectral camera, plus an individual synthetic blue band. The described procedure for dataset construction can be replicated for crops other than sugar cane, or be used to generate new data for other semantic segmentation purposes. Remote sensing and data science provide a technology framework applicable to farm management, sustainability and sustainability of agriculture are two of the most important aspects of farmers' decision making. The most relevant contributions of this project are:

- Making sugar cane crop datasets available for field data analysis and decision making.

- The use of geographic information system tools to increase the resolution of orthomosaics through interpolation.

- The design of a multivariate linear regression model to create a representative orthomosaic of the blue band.

References

Arenas, R. (2016). Aplicación de la detección en la exploración geominera y de recursos naturales. Oviedo: Universidad Oviedo.

Arivazhagan, S. e. (2013). Ingeniería Agrícola Internacional. *CIGR Journal*, 211-217.

Castillejo-González, I., López-Granados, F., García-Ferrer, A., Peña-Barragán, J., Jurado-Expósito, M., Sánchez de la Orden, M., & González-Audicana, M. (2009). Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Computers and Electronics in Agriculture*, 207–215.

Di Cicco, M., Potena, C., Grisetti , G., & Pretto, A. (2017). Automatic Model Based Dataset Generation for Fast and Accurate Crop and Weeds Detection. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5188-5195.

Donyavi, Z., & Asadi, S. (2020). Diverse training dataset generation based on a multi-objective optimization for semi-Supervised classification. *Pattern Recognition*, 16.

Figueredo-Ávila, G. A., & Ballesteros-Ricaurte, J. A. (2016). Identificación del estado de madurez de las frutas con redes neuronales artificiales, una revisión. *Ciencia y Agricultura*, 117-132.

Foody, G. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 185–201.

Gašparović, M., Zrinjski, M., Barković, D., & Radočaj, D. (2020). An automatic method for weed mapping in oat fields based on UAV imagery. *Computers and Electronics in Agriculture*, 1-12.

Grupta, S. D., & Ibaraki, Y. (2014). *Applications of RGB color imaging in plants.* Boca Ratón: CRC Press.

Guarneri, J., & Weih, R. J. (2010). Comparing Methods for Interpolation to Improve Raster Digital Comparing Methods for Interpolation to Improve Raster Digital Elevation Models Elevation Models. *Journal of Physics: Conference Series*, 77-81.

Hu, L., Luo, X., & Wei, Y. (2020). Hyperspectral Image Classification of Convolutional Neural Network Combined with Valuable Samples. *Journal of Physics: Conference Series*, 8.

Kim, M., Lee, J., Han, D., Shin, M., & Im, J. (2018). Convolutional Neural Network-Based Land Cover Classification Using 2-D Spectral Reflectance Curve Graphs With Multitemporal Satellite Imagery. *IEEE Journal Of Selected Topics in Applied Earth Observations and Remote Sensing*, 4604-4617.

Lee, Y., Son, J., & Kim, T. (2022). Determination of Optimal Ground Sampling Distance for Matching GCP Chips and Satellite Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 287-292.

Lei, L. X., & Peng, L. (2020). Training strategy of CNN for remote sensing image classification with active learning. *IOP Conference Series: Earth and Environmental Science*, 1-9.

Li, B., Hua, Y., & Lu, M. (2021). Advanced Multiple Linear Regression Based Dark Channel Prior Applied on Dehazing Image and Generating Synthetic Haze. *Advances in Science, Technology and Engineering Systems Journal*, 1-10.

Lillesand, T., & Kiefer, R. &. (2007). *Remote sensing and image interpretation.* U.S.A: .: Jhon Wiley & Sons Inc. doi: 978-0-47-005245-7

Ling, M., Cheng, Q., Peng, J., Zhai, C., & Jiang, L. (2022). Image Semantic Segmentation Method Based on Deep Learning in UAV Aerial Remote Sensing Image. *Mathematical Problems in Engineering*, 1-10.

Lira, J. (2003). *La percepción remota: nuestros ojos desde el espacio* (Tercera ed.). México: FCE, SEP, Conacyt. doi:9681669223

Lyu , Y., Vosselman , G., Xia , G.-S., & Yilm, A. (2020). UAVid: A semantic segmentation dataset for UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108-119.

ISSN 2531-2960

Mahlein, A. K. (2016). Plant disease detection by imaging sensors - Parallel and specific demand for precision agriculture and plant phenotyping. *Plant disease*, 241-251.

Olsson, P.-O., Vivekar, A., Adler, K., García, V., Koc, A., Alamrani, M., & Eklundh, L. (2021). Radiometric Correction of Multispectral UAS Images: Evaluating the Accuracy of the Parrot Sequoia Camera and Sunshine Sensor. *Remote Sensing*, 577.

Tripicchio , P., Satler , M., Dabisias, G., Ruffald, E., & Avizzano, C. (2015). Towards Smart Farming and Sustainable Agriculture with Drones. *2015 International Conference on Intelligent Environments*, 140-143.

Vassilopoulou, S., Hurni, L., & Dietrich, V. (2002). Orthophoto generation using IKONOS imagery and high-resolution DEM: a case study on volcanic hazard monitoring of Nisyros Island (Greece). *ISPRS Journal of Photogrammetry & Remote Sensing*, 24-38.

Zabawa, L., Kicherer, A., Klingbeil, L., Opter, R. T., Kuhlmann, H., & Roscher, R. (2020). Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 73–83.