Anuário do Instituto de Geociências
Universidade Federal do Rio de Janeiro
https://revistas.ufrj.br/index.php/aigeo/

ISSN 0101-9759
e-ISSN 1982-3908

Development of a Low-Cost Terrestrial Mobile Mapping System for Urban Vegetation Detection Using Convolutional Neural Networks

Desenvolvimento de um Sistema de Mapeamento Móvel Terrestre de Baixo Custo para Detecção da Vegetação Urbana Usando Redes Neurais Convolucionais

Kauê de Moraes Vestena[1] & Daniel Rodrigues dos Santos[2]

[1] Universidade Federal do Paraná, Departamento de Geomática, Curitiba, PR, Brasil
[2] Instituto Militar de Engenharia, Seção de Engenharia Cartográfica, Rio de Janeiro, RJ, Brasil
Corresponding author: Kauê de Moraes Vestena; kauemv2@gmail.com

Abstract

Urbanization brought a lot of pollution-related issues that are mitigable by the presence of urban vegetation. Therefore, it is necessary to map vegetation in urban areas, to assist the planning and implementation of public policies. As a technology presented in the last decades, the so-called Terrestrial Mobile Mapping Systems - TMMS, are capable of providing cost and time effective data acquisition, they are composed primarily by a Navigation System and an Imaging System, both mounted on a rigid platform, attachable to the top of a ground vehicle. In this context, it is proposed the creation of a low-cost TMMS, which has the feature of imaging in the near-infrared (NIR) where the vegetation is highly discriminable. After the image acquisition step, it becomes necessary for the semantic segmentation of vegetation and non-vegetation. The current state of the art algorithms in semantic segmentation scope are the Convolutional Neural Networks - CNNs. In this study, CNNs were trained and tested, reaching a mean value of 83% for the Intersection Over Union (IoU) indicator. From the results obtained, which demonstrated good performance for the trained neural network, it is possible to conclude that the developed TMMS is suitable to capture data regarding urban vegetation.

Keywords: Mobile geospatial data acquisition systems; NIR Imaging; Semantic segmentation

Resumo

A urbanização acarretou muitas problemáticas relacionadas com a poluição, mitigáveis pela presença de vegetação urbana. Por conseguinte, é necessário mapear a vegetação nas áreas urbanas, de modo a apoiar o planejamento e a implementação de políticas públicas. Como tecnologia apresentada nas últimas décadas, os denominados Sistemas de Cartografia Móvel Terrestre - SMMT, capazes de proporcionar uma aquisição de dados eficaz em termos de custo e tempo, são compostos principalmente por um Sistema de Navegação e um Sistema de imageamento, ambos montados sobre uma plataforma rígida, fixáveis à parte superior de um veículo terrestre. Neste contexto, propõe-se a criação de um SMMT de baixo custo, dotado da capacidade de gerar imagens contendo o infravermelho próximo (NIR), onde a vegetação é altamente discriminável. Após a etapa de aquisição das imagens, se faz necessária a segmentação semântica da vegetação e a não-vegetação. Os atuais algoritmos de estado da arte no âmbito da segmentação semântica são as Redes Neurais Convolucionais - CNNs. Neste estudo, as CNNs foram treinadas e testadas, atingindo um valor médio de 83% para o indicador Intersecção sobre União (IoU). A partir dos resultados obtidos, que demonstraram bom desempenho para a rede neural treinada, é possível concluir que o TMMS desenvolvido é adequado para captar dados relativos à vegetação urbana.

Palavras-chave: Sistemas móveis de aquisição de dados geoespaciais; Imageamento NIR; Segmentação semântica

Received: 29 August 2021; Accepted: 30 November 2021

CC BY

1 Introduction

Nowadays, the outcomes of the urbanization process are increasingly more visible. There are a myriad of issues related to this process that have direct influence on the urban population quality of life, like all kinds of pollution. Many of these problems can be mitigated by a solid presence of vegetation in the urban scenarios, accordingly to Nicodemo & Primavesi (2009) , the major notably gains provided by them are: 1) Microclimate Control, by the alleviation of the local temperature, reducing the occurrence of heat islands; 2) Air Pollution reduction, with the surface of leaves absorbing some of the air pollutants and holding some of the biggest solid particles; 3) Noise Reduction, the leaves acts as barriers against the sound waves, and also grass is a good pavement against sound propagation; 4) Rainfall interception, the process of the growing of the plant roots and the deposition of organic material, both contribute to improve the urban soil permeability; 5) Carbon Retention, by the natural plant breathing process, that absorbs more carbon dioxide than delivers to the atmosphere; 6) cultural and aesthetic values, some places prizes the presence of vegetation, e.g. by legal fostering and tax reductions.

Despite all of these benefits, urban vegetation can cause damage if grown under improper conditions, so both faces are sources for purposing the mapping activity of this critical matter. The demand for this kind of data is a very challenging one, as a regular city can accommodate several kilometers of roads, setting up a scenery that is susceptible to quick changes along the years. So, the mapping technique must fulfill the feature of quick data capture of an entire study area (that can be an entire city).

According to El-Sheimy (2005), the so-called Terrestrial Mobile Mapping Systems - TMMS can attend to that demand. This kind of system should contain four essential subsystems: an imaging system; a navigation system; a control, storage and power supply system; and a rigid platform. The TMMS can exist in many shapes and configurations, and there are a lot of possible choices for their components that can exist in every range of monetary cost. TMMS can rely on active and/or passive imaging systems, but generally only photographic cameras (passive sensors) are viable for a low-budget setup. Low-cost systems are generally desirable, as they allow for multiple agents to act in the field, in a way that can reduce largely the demand of time to accomplish a survey. And also, capturing data in Near-Infrared (NIR) spectrum is an additional feature to acquire relevant information about vegetation (Myneni et al. 1995).

Imagery is a feature-rich kind of data, especially in urban scenes, they can have a lot of other themes as well as vegetation: pavement, cars, people, buildings, the sky, etc. So, for any later activity aiming for the vegetation mapping is mandatory the segmentation and labeling (semantic segmentation) of every pixel in each photograph into vegetation or the other features. Nowadays, in line with many authors, like (Chen et al. 2018; Mostjabi, Yadollahpour & Shakhnarovich 2015; Badrinarayanan, Kendall & Cipolla 2017). The best kind of algorithms for semantic segmentation are the ones based on Convolutional Neural Networks - CNNs, that are networks trained by Deep Learning algorithms, this means that no explicit directions are given to it in order to accomplish its goals, actually, only examples of correct classifications are given, then the network will iteratively adjust its weights to improve their own performance.

In order to achieve those goals, the present paper presents the development of a low-cost TMMS capable of capturing the NIR spectrum and the use of CNNs for the segmentation process of the acquired imagery. The main steps in this work are: the proposal and selection of the TMMS components; establish the data collection and processing protocols; train a CNN capable to carry out the semantic segmentation process; evaluate the accuracy of the semantic segmentation, by cross-checking the obtained results with ground-truth samples. The highlights of the proposed method are:

- The development of a Low-Cost TMMS, capable of IR imaging; and
- The employment of the state-of-art algorithm for semantic segmentation.

2 Material and Method

2.1 Materials and Resources

In order to develop the present piece of work, a bunch of devices and resources that are physical, computational or even procedural have been employed. The process of resource selection has been carried out by some guidelines: the low cost; the availability; and mandatory preference for open-sourced materials.

The TMMS itself was built upon a rigid platform that is made of medium-density fiberboard painted in white. On its “bottom side” there are two bars with universal suction couplers. In its “upper side” there are the functional devices of the system, sorted here by the relevance of each one: A “Raspicam V2” camera module with 8MP resolution with infrared capturing capacity, serving as the exteroceptive (imaging) sensor; An Adafruit BNO055

Inertial Measurement Unit (IMU); An GNSS-UBX-M8030 module, capable of capturing data up to 3 constellations simultaneously, that alongside with the IMU compounds the positioning system of the TMMS; A Raspberry Pi 3B board, the computer of the platform, that acts both as the hardware-level integrator of the imaging and positioning systems, power-sourcing them and providing a unified time scale for data capture, and also as a software-level integrator, running the sensors drivers and the data recording and preprocessing software. Besides the functional part, there are other essential pieces that are necessary for the TMMS working, such as the 12 V battery used as power-source, the SD Card used as storage and a red-light Filter that acts blocking the green and blue part of the visible spectrum. The platform, which is presented in Figure 1, had its actual cost at around BRL 1200.

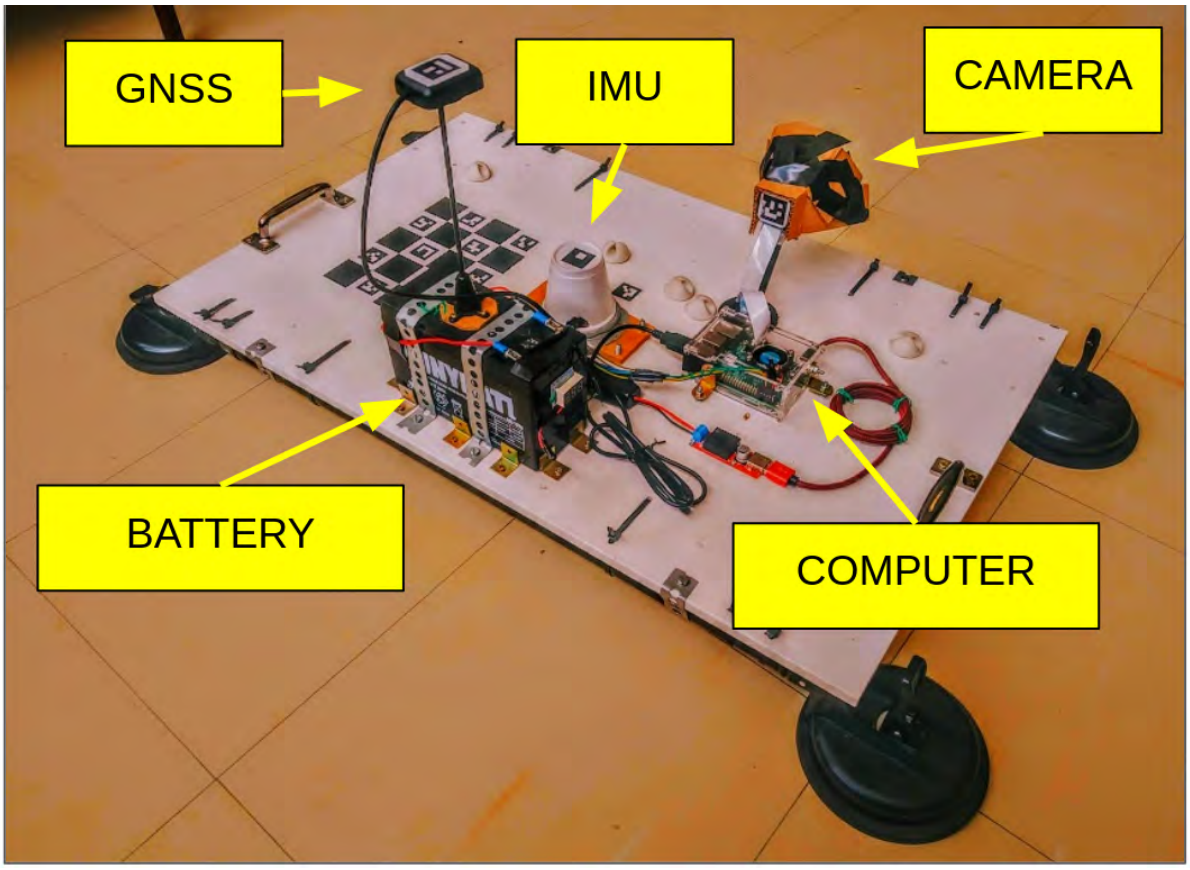

Figure 1 The developed TMMS platform.

Also, in order to handle the data capture and let the user of the platform take control of this process, there are the software stuff running on the embedded computer: Lubuntu 16.04 as the operating system; the Robot Operating System (ROS) as the framework interfacing the equipment along with the realization of the time system and the creation of the data files; and the Real VNC (Virtual Network Computing) implementation of the VNC protocol as a server for remote controlling.

Outside the platform, in order to use a CNN architecture in order to train, evaluate and use the neural networks, Google's Tensorflow library has been used, for statistics, the functions from the *SciPy* library have been used. All the codes used for data processing are available at the repository: https://github.com/kauevestena/snav.

2.2 Methodology

The proposed method is divided in two main stages: data acquisition and data processing. The preliminary stage covers the entire process for training and evaluation of the CNN. After, it starts the planning of a surveying, passing to its execution. Finally, the data are processed and the geotagged vegetation-only images are obtained. Figure 2 presents the workflow of the proposed method.

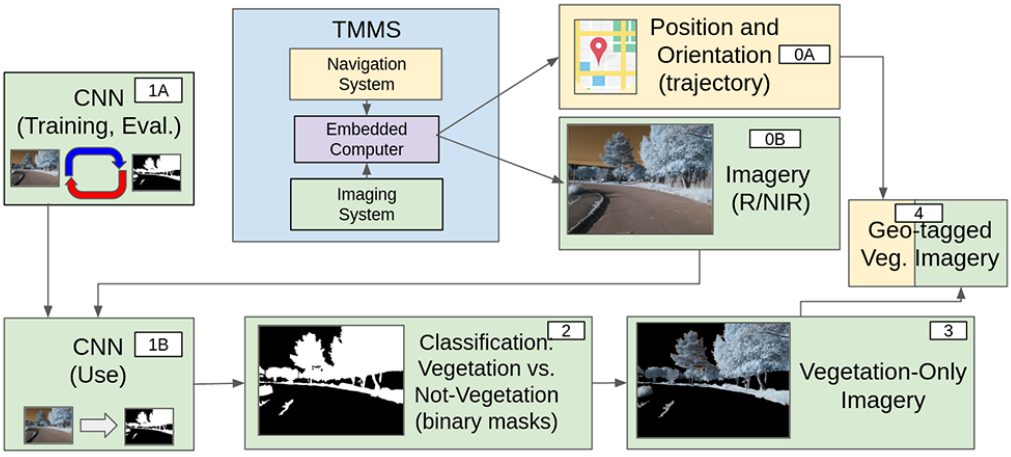

Figure 2 The workflow: 0 is data acquisition; 1 is the use of the CNN; 2 to 4 is the remaining part aiming to the final product.

For the data acquisition stage, there are two parts worth considering: data recording and data collection. The first refers to how the system as a whole will spawn the data files, aside from its purpose. The latter refers to the surveying execution process itself.

The data recording was made based on the ROS Kinetic, all the data is stored in a file that is made by it. So, each equipment in the TMMS will generate its own data: the camera generates RED/NIR images at a rate up to 90 Hz and a resolution up to 8 MP; the IMU captures at up to 100 Hz the unitary quaternion that magnetic-north oriented absolute orientation, the free linear acceleration and the angular velocity; and the GNSS receiver will record the position as geodetic coordinates and the velocity as an east-north-up vector. The control schema, which is depicted in Figure 3, is done by running the Real VNC server at the computer, which allows one to put a remote device (generally a smartphone, connected to the same network) in charge of the record process, both to start and to stop it. Once the recording process finishes, we have the data file in a binary format.

The data collection part depends on some planning, with the route to be ran as the main part. Three factors must be taken in account for route planning: the area of interest; the coverage, that is ensuring that the planned route will attend to supply all the necessary data; and the viability, which means that the planned route does not violate any traffic law. Figure 4 shows the platform mounted in a car.

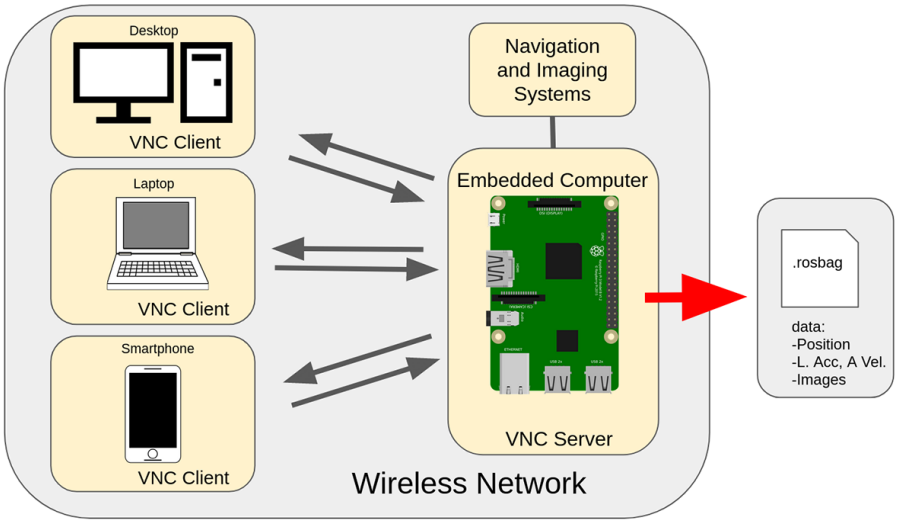

Figure 3 The Platform Control Scheme.

Figure 4 The Platform mounted in a car. Detail: remote controlling the System.

After the data collection, we extract the images and positioning data from the binary file, then we process them separately. In order to align both the positional data and the images within the time domain, we have used interpolation: for the GNSS data we used linear interpolation, since it records data at 1 Hz, we also first linearized the geodetic coordinates to ENU (East North Up) vector coordinates, interpolated, to then go back to geodetic coordinates; for the IMU data, we have employed the nearest neighbor, since it generally operates at 100 Hz.

In Figure 5, the digital images are used for the training of the CNN, and for the application of the CNN. As samples for the CNN training, we have classified manually a total of 50 classification masks (binary images, where white is vegetation), subdivided in three categories: I) "Train", the images used to effectively train the CNN; II) "Validation", the images used as internal validation by the CNN, and therefore for the fine tuning of the CNN parameters; and III) "Test", that are images not used by the training algorithm at all, which finality is to externally evaluate the quality of the classification done by the CNN. As a model of CNN. We have opted for the full-resolution residual networks (FRRN) (Pohlen et al. 2017), since it was designed for street scenes. They make use of skip connections with both full and reduced resolution images.

The CNN training was done along hundreds of cycles named epochs, as shown in Figure 6, in which the entire dataset is passed through the CNN. The images are downsampled to the resolution of 512 x 512, such as suggested by Sabotke & Spieler (2020) and Kannojia & Jaiswal (2018), since the full resolution causes more computational burden than benefit in terms of accuracy. In order to input a random factor to mitigate the effects of the sensor position and its particular sensibility for the light, a random brightness variation of 10% as an image augmentation technique was used (Mikołajczyk & Grochowski 2018). The performance of the CNN is continuously accessed by some error metrics, at each epoch they are evaluated internally using the imagery of group II. At the early epochs, the CNN will perform poorly, then will continuously raise, until it reaches a certain plateau, from then it would not improve further, hence the training can be stopped.

The employed error metrics are based on a statistical confusion-matrix (Fawcett 2006), comparing for each pixel respectively the ground truth with the classification done by the CNN: V (vegetation) and V is a True Positive (TP); NV (non-vegetation) and NV is a True Negative (TN); V and NV is a False Negative; NV and V is a False Positive. In the binary classification of this work, we consider the TP as much more meaningful hits, since in the R-NIR images there are much more features distinguishable from vegetation than the opposite. Furthermore, the FP are considered as the worst error, as one can run analysis in the "vegetation-only" imagery that was meant to be run in vegetation pixels only.

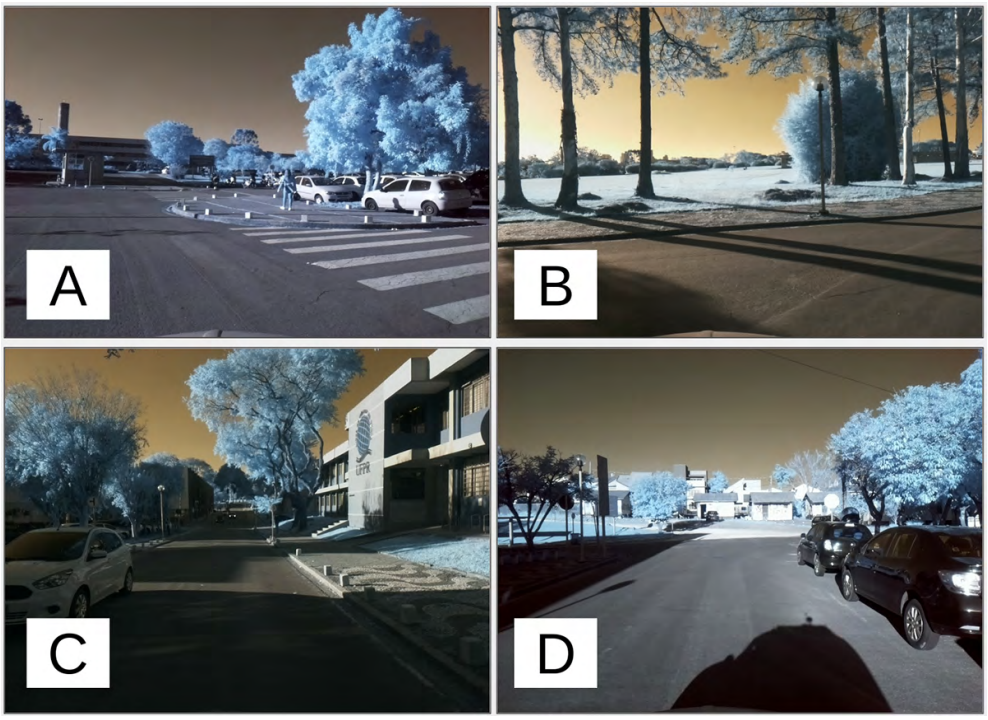

Figure 5 Samples of digital images taken by the TMMS.

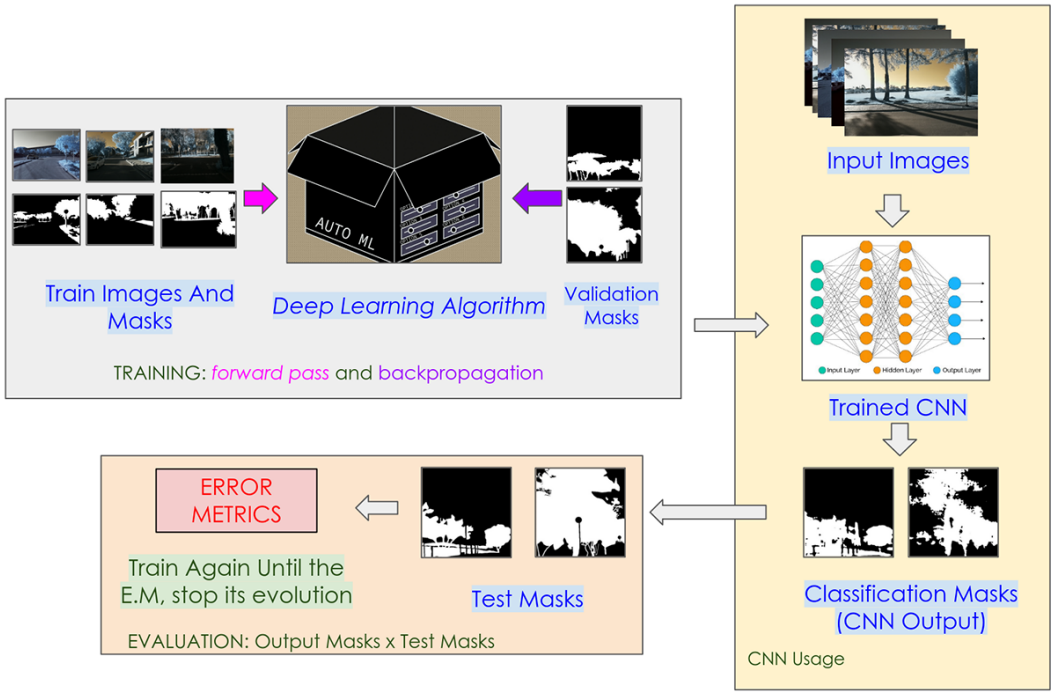

Figure 6 The CNN Training Process.

The total of pixels for each of the four possibilities is added, and the five employed metrics (Fawcett 2006) is computed: Hit Rate (HR) that is the rate between hits and total pixels; Positive Predictive Value (PPV) that is the rate between TP and the pixels classified as vegetation; Sensitivity (S) that is the rate between TP and pixels that are actually vegetation; F1-Score (F1) that is the harmonic mean between S and PPV; and Intersection Over Union (IoU) that is the ratio between the intersection of vegetation predictions and vegetation ground truth (that are the TP) and the union of the same groups. All those metrics are formulated as follows, in the Equations 1-5, respectively.

$$HR = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

$$PPV = TP / (TP + FP) \quad (2)$$

$$S = TP / (TP + FN) \quad (3)$$

$$F1 = (PPV * S) / (PPV + S) = 2TP / (2*TP + FP + FN) \quad (4)$$

$$IOU = TP / (TP + FP + FN) \quad (5)$$

Thus, the mean value of any metric for each epoch is calculated. In this work, the best epoch was chosen as the one who achieved the best average IOU. After the algorithm chooses the best epoch one can use it as the CNN to do the classification for any image taken by the TMMS.

3 Experiments

3.1 Experiment Design

The experiments taken on the behalf of this work were two: 1) a complete survey with the TMMS; and 2) training and validation of many epochs for the CNN.

For 1), the parameters employed for the capture are the following: recording at 100, 1 and 10 Hz for the IMU, GNSS and camera respectively; GPS, GLONASS and GALILEO are the employed GNSS constellations; the "automotive mode" (non-holonomic constraints) was activated for the GNSS; the images were recorded at a 1280x960 pixels of resolution with an exposure time of 1/1000 s. In a route of 2.1 km traversed in 11.9 minutes, depicted in Figure 7, 7111 photos were taken, with 4.6 GB of uncompressed data generated.

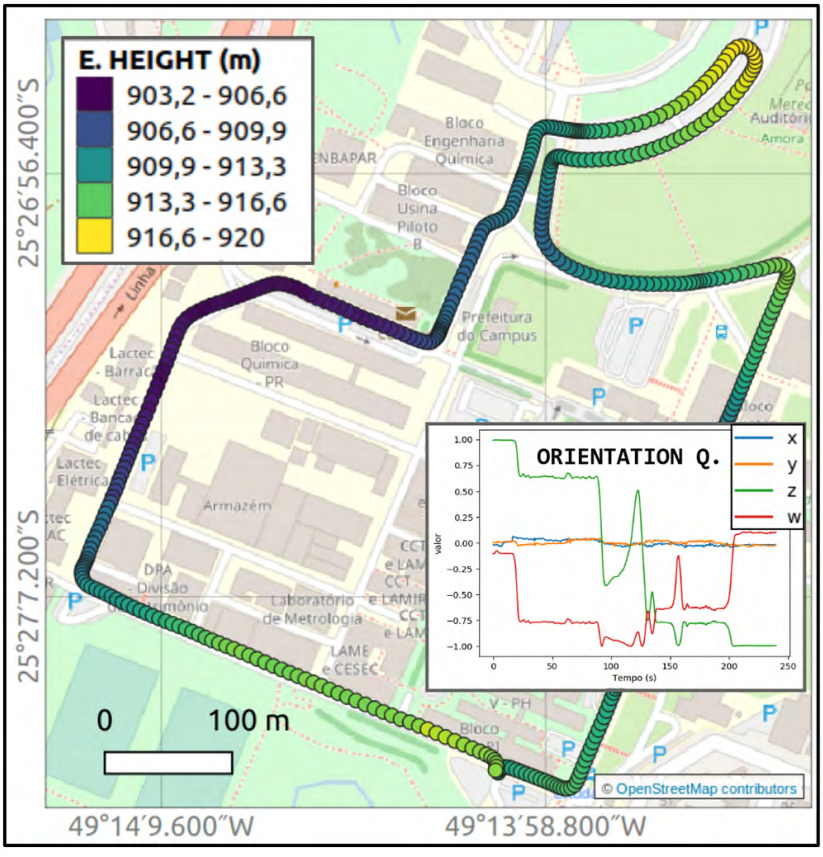

Figure 7 Part of the survey route, with the orientation quaternion along in detail.

Then, for 2), we classified considering every feature with leaves as vegetation (including grass); the imagery was downsized to 512x512 pixels; 1100 epochs as the stop criteria; 40 images as the size of the dataset, for the training that was done with a batch size of 20 pixels; the FRRN (Pohlen et al. 2017) with a InceptionV4 frontend are the model for CNN. We employed Google's Colab Servers with GPU acceleration for the training, each epoch was trained in approximately 5 min, taking almost 4 days to train.

3.2 Results and Analysis

Firstly, regarding the positioning data, no numerical analysis has been carried out, but, as shown in Figure 7, the collected data suited finely to an existing map data, totally independent from the TMMS data, showing its consistency both in an absolute manner as well in a relative aspect (appropriate straight and curved paths without jumps). The orientation data also performed well, with variations mostly on "z" and "w" components of the orientation quaternion, showing a correctly level-oriented system.

For the CNN data, in Figure 8 are depicted the detailed IoU evolution for the CNN training, with the mean value for the test dataset per epoch, along with min, max values and the positive and negative interval including the standard deviation.

Figure 8 shows the expected evolution for the CNN performance, with a high gain in the initial epochs, then it progressively slows down until it stops evolving (hence vulnerable to overparameterization), so the training can be stopped. The epoch numbered 1059 was chosen as the best epoch, since it achieved a IoU mean ratio of 0.8397, the best one among the data.

In Figure 9 the boxplot with the summary of all the employed error metrics are depicted.

As shown in Figure 9, the HR metric overestimates the CNN quality, as it considers the TN values that can be background and are less relevant in this binary classification as stated in Chapter II. Regarding the PPV and Sensibility as complementary, as the CNN performed better in the latter, we can see that the CNN commits less FN as it weighs less compared to the TP. Comparing the IoU with the F1 index, the latter gives too much weight to the FPs, overestimating the CNN quality. So, the IoU was chosen as the best metric to meaningfully summarize the CNN performance, with 0.8397 as mean score with 0.0217 of standard-error.

As this work is intending to provide geotagged vegetation-only R-IR imagery as the main goal, in Figure 10 a sample is presented.

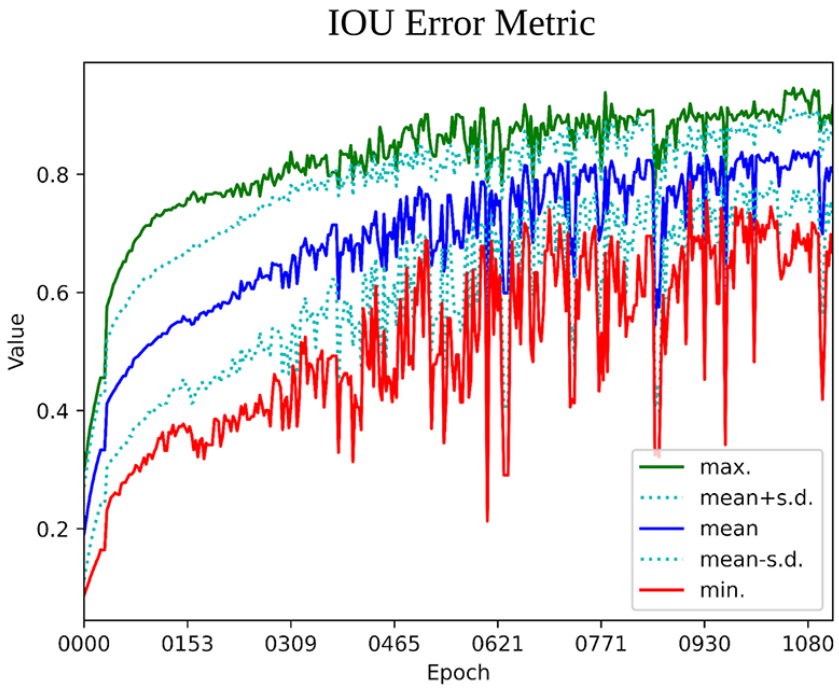

Figure 8 The evolution of the IoU error metrics along the epochs.

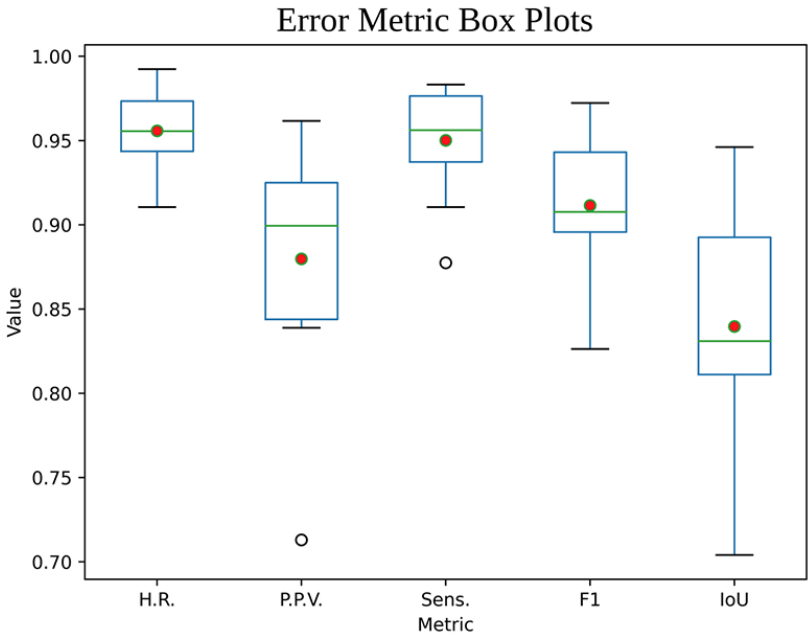

Figure 9 The boxplot with the summary of the error metrics with mean in red.

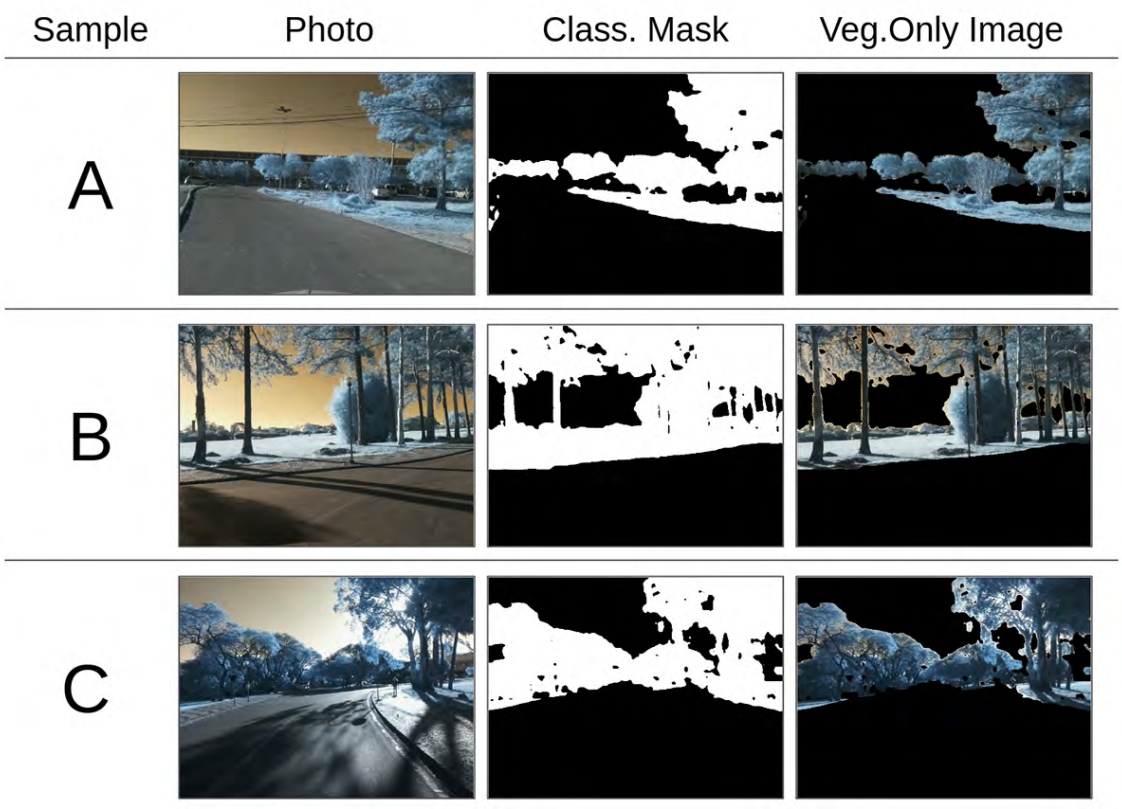

Figure 10 Samples of resulting vegetation-only images.

As a final result, a sample map has been generated with the experiment data, with a simple approach: the degree of incidence of vegetation is shown as the percentage of vegetation pixels. The complete version (in Brazilian Portuguese) of the generated map is available at the "map" folder of the GitHub repository presented on Chapter II, a preview of it is rendered at Figure 11.

4 Conclusions and Future Works

This paper presented two pertinent themes concerning the geosciences area: a development of a complete mobile mapping system; and the classification of its imagery as a preparatory step in order to use them for urban vegetation studies, along with the positioning and orientation data.

The assembly of the TMMS platform was a success and the system is fully operational and can be used to collect data, as well as its project is fully open at the author's master thesis, in Portuguese, available at: https://rebrand.ly/dissertacao_kaue. So anyone with the technical knowledge can replicate the hardware part as well as the software part which are available on GitHub. The platform design and the project as a whole has successfully reached its goal as a low-cost purpose.

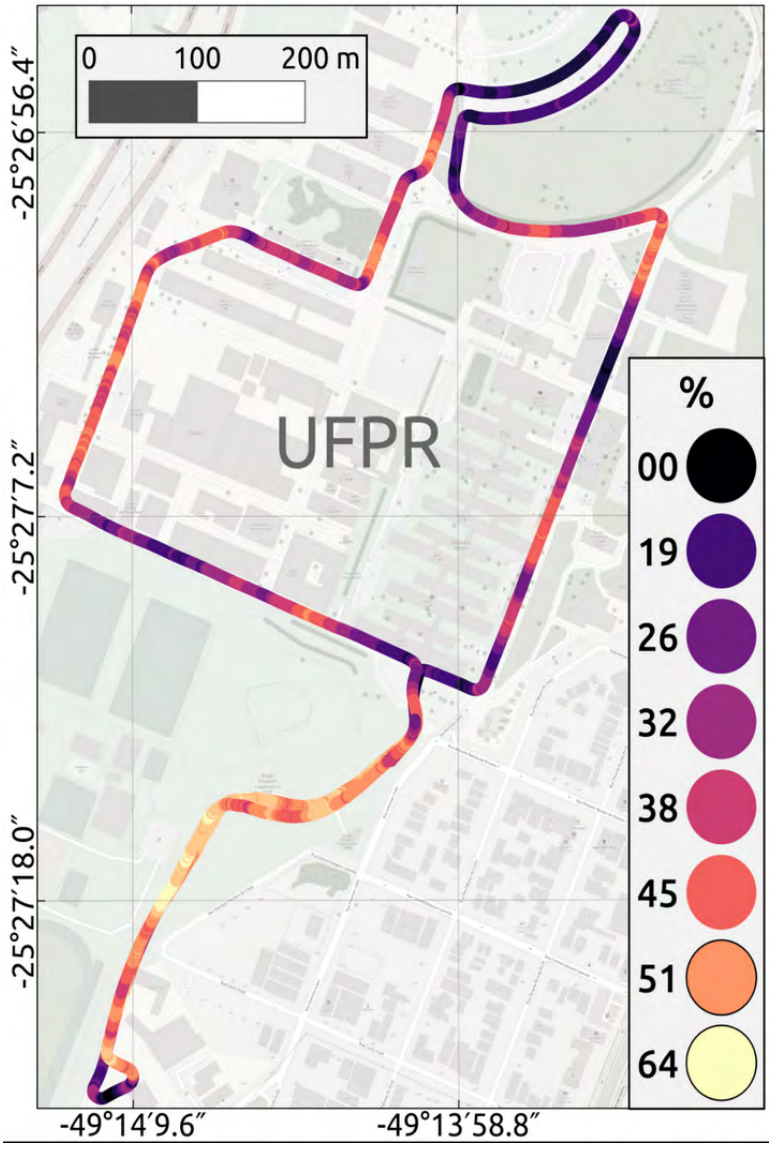

Figure 11 Preview of the generated sample map.

The presented results are satisfactory, then it can be said that the project objectives have been fulfilled. The TMMS have potential to be employed in a large-scale data collection schema, as it can be replicated and be embedded in several vehicles in order to map an entire big city. The collected data has a huge vocation to be used in urban vegetation studies, including vegetation health and presence, with high-resolution data, since the camera capture imagery notably close from the subjects, although this strong point is also its most notorious limitation: this imagery only shows vegetation attached to streets, one cannot see inside city blocks without help from aerial/orbital imagery, aside from really tall trees.

As future work, the following are suggested: to train specific CNN to identify shadows, tree trunk, tree foliage, tree pathologies, etc.; to develop a tree counting algorithm, to calculate the 3D coordinate of the tree base; to employ a luminance sensor to adapt sensor gains accordingly to light changes; to use many epochs in order to create a voting schema, to improve classification; and to integrate more geometric constraints.

5 Acknowledgements

This project was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES.

6 References

Badrinarayanan, V., Kendall, A. & Cipolla, R. 2017, 'Segnet: A deep convolutional encoder-decoder architecture for image segmentation', *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481-95. https://doi.org/10.1109/TPAMI.2016.2644615

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A.L. 2018, 'Deep Lab: Semantic image segmentation with deep convolutional nets and fully connected CRFS', *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 4, pp. 834-48. https://doi.org/10.1109/TPAMI.2017.2699184

El-Sheimy, N. 2005, 'An overview of mobile mapping systems', *Proceedings of the FIG Working Week*, pp. 16-21.

Fawcett, T. 2006, 'An introduction to ROC analysis', *Pattern recognition letters*, vol. 27, no. 8, pp. 861-74. https://doi.org/10.1016/j.patrec.2005.10.010

Kannojia, S.P. & Jaiswal, G. 2018, 'Effects of varying resolution on performance of CNN based image classification: An experimental study', *International Journal of Computer Sciences and Engineering*, vol. 6, no. 9, pp. 451-6. http://dx.doi.org/10.26438/ijcse/v6i9.451456

Mikołajczyk, A. & Grochowski, M. 2018, 'Data augmentation for improving deep learning in image classification problem', *2018 International Interdisciplinary PhD workshop (IIPhDW)*, pp. 117-22, IEEE. https://doi.org/10.1109/IIPHDW.2018.8388338

Mostajabi, M., Yadollahpour, P. & Shakhnarovich, G. 2015*,* 'Feedforward semantic segmentat*ion* with zoom-out features'*,* *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 3376-85. https://doi.org/10.48550/arXiv.1412.0774

Myneni, R.B., Hall, F.G., Sellers, P.J. & Marshak, A.L. 1995, 'The interpretation of spectral vegetation indexes', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 2, pp. 481-6. https://doi.org/10.1109/TGRS.1995.8746029

Nicodemo, M.L.F. & Primavesi, O. 2009, 'Por que manter árvores na área urbana?', *Embrapa Pecuária Sudeste-Documentos (INFOTECA-E)*.

Pohlen, T., Hermans, A., Mathias, M. & Leibe, B. 2017*,* 'Full-resolution residual networks for semantic segmentation in street scenes'*,* *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 4151-60. https://doi.org/10.48550/arXiv.1611.08323

Sabottke, C.F. & Spieler, B.M. 2020, 'The effect of image resolution on deep learning in radiography', *Radiology: Artificial Intelligence*, vol. 2, no. 1, pp. e190015. https://doi.org/10.1148/ryai.2019190015

Author contributions
Both authors contributed equally to this work.

Conflict of interest
The authors declare no potential conflict of interest.

Data availability statement
Model data are freely available on https://github.com/kauevestena/snav Reference datasets can be downloaded from: https://github.com/kauevestena/snav/tree/master/dataset
Scripts and code are available freely under MIT license
All data included in this study are publicly available in the literature.

Funding information
This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Editor-in-chief
Dr. Claudine Dereczynski

Associate Editor
Dr. Silvio Roberto de Oliveira Filho.

How to cite:
Vestena, K.M. & Santos, D.R. 2022, 'Development of a Low-Cost Terrestrial Mobile Mapping System for Urban Vegetation Detection Using Convolutional Neural Networks', *Anuário do Instituto de Geociências*, 45:46008. https://doi.org/10.11137/1982-3908_45_46008