Inteligencia Artificial 24(68) (2021), 89-103
doi: 10.4114/intartif.vol24iss68pp89-103

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Performance Analysis in the Segmentation of urban paved roads in RGB satellite images using K-Means++ and SegNet: case study in São Luís-MA

João Batista Pacheco Junior[1], Henrique Mariano Costa do Amaral[2]

[1]Universidade Estadual do Maranhão, Maranhão, Brasil
ioannis.baptista@gmail.com

[2]Universidade Estadual do Maranhão, Maranhão, Brasil
hmca13@gmail.com

Abstract The design and manual insertion of new terrestrial roads into geographic databases is a frequent activity in geoprocessing and their demand usually occurs as the most up-to-date satellite imagery of the territory is acquired. Continually, new urban and rural occupations emerge, for which specific vector geometries need to be designed to characterize the cartographic inputs and accommodate the relevant associated data. Therefore, it is convenient to develop a computational tool that, with the help of artificial intelligence, automates what is possible in this respect, since manual editing depends on the limits of user agility, and does it in images that are usually easy and free to access.

To test the feasibility of this proposal, a database of RGB images containing asphalted urban roads is presented to the K-Means++ algorithm and the SegNet Convolutional Neural Network, and the performance of each one was evaluated and compared for accuracy and IoU of road identification.

Under the conditions of the experiment, K-Means++ achieved poor and unviable results for use in a real-life application involving asphalt road detection in RGB satellite images, with average accuracy ranging from 41.67% to 64.19% and average IoU of 12.30% to 16.16%, depending on the preprocessing strategy used. On the other hand, the SegNet Convolutional Neural Network proved to be appropriate for precision applications not sensitive to discontinuities, achieving an average accuracy of 87.12% and an average IoU of 71.93%.

Keywords: Geoprocessing, RGB image, image segmentation, K-Means++, Convolutional Neural Network.

1 Introduction

The booming use of Geographic Information Systems (GIS) for many activities related to geoprocessing requires an increasing level of quality of geospatial data. Extracting such information from remotely detected images is a practical method to feed geographic databases, but the georeferencing of this information requires, in many cases, visual detection by a human operator and the manual vectorization of the geometries that guarantee such georeferencing. This process requires considerable time and labor, given the usual volume of elements present in the images.

One of the processes done manually in many geoprocessing jobs is the incorporation of new land routes (i.e., streets, avenues, highways, roads, alleys, etc.) into geographic databases, as more up-to-date satellite images of the territory are acquired. New urban (neighborhoods, subnormal agglomerations, isolated urban areas, etc.) and rural (village, hamlets, settlement projects, etc.) occupations continually emerge, for which specific vector geometries need to be designed to characterize the cartographic inputs (in particular, streets) and accommodation of relevant associated data. This activity requires the user to analyze several areas of the territory covered by the image, and update the geometry of the roads according to the detected changes, conditioning this process to the limits of human agility. With this in mind, an increase in productivity is expected, having in hand tools that automate whatever is possible in such a process, using available resources, such as aerial images of the territory and computational intelligence.

With regard to the use of images, the multispectral type, which has a thermal band, has been a tool that facilitates the differentiation and classification of objects in remote sensing images. As demonstrated by Nóbrega [1] in his study, reflectance can be used as a determining attribute in detecting discontinuities between different objects in the image, as illustrated and exemplified in Figure 1.

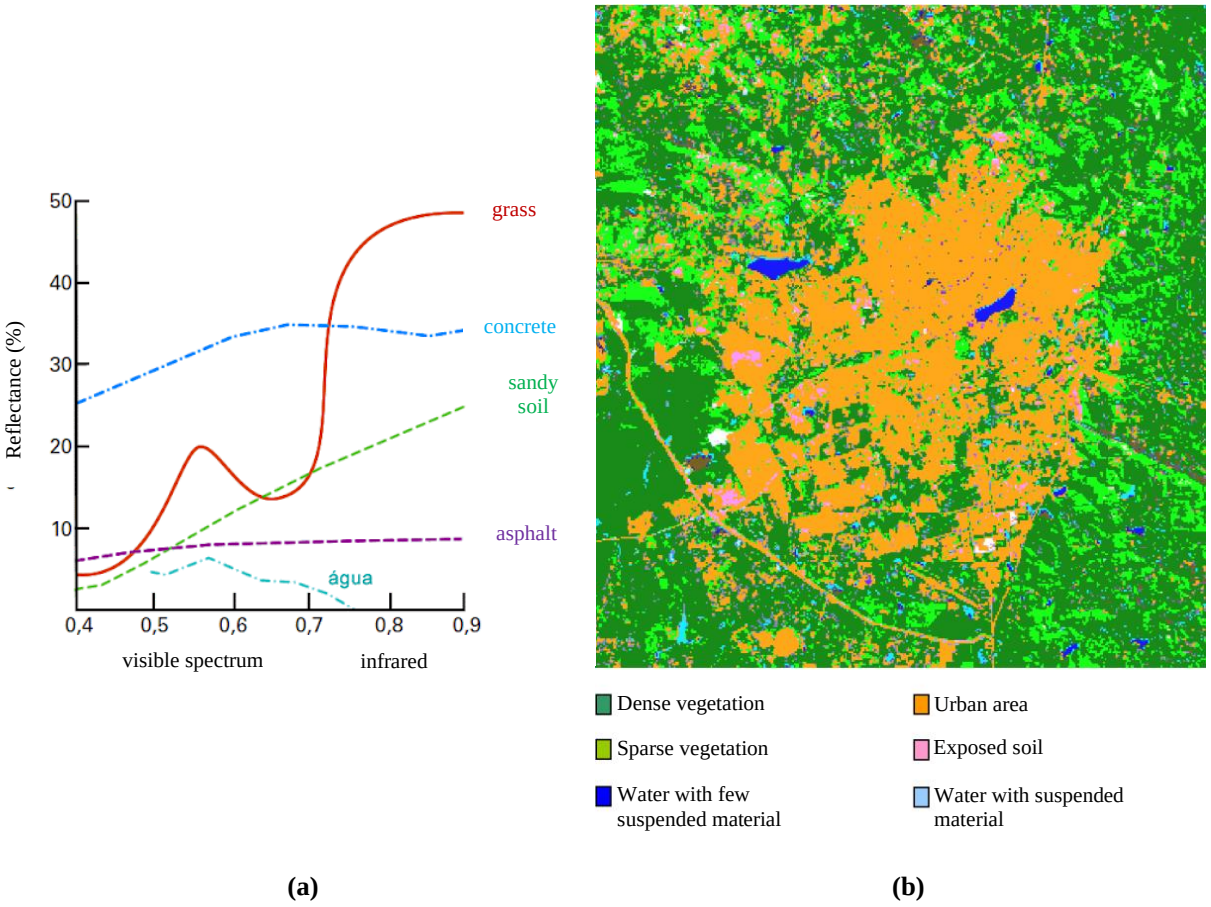


Figure 1. a) Percentage of reflectance for each wavelength value (μm) of the most common classes in urban landscapes [2]. Its differentiation in the infrared spectrum is evident (> 0.7), which is not present in most free geoservices. b) Example of multispectral image (TM/LANDSAT-5), classified with thermal band facilities [3].

However, multispectral images with high spatial resolution are expensive inputs and, therefore, difficult to acquire. Thus, lies in this aspect the great advantage of widely used geoservices: their free and easy access. This makes it common for most of the vector editing work to be done manually, based on images from free and widely used geoservices (such as Google and Bing). This entails, therefore, the development of a computational technique capable of satisfactorily differentiating objects in RGB satellite images, in order to compensate for the absence of the thermal band.

Given the context presented, this experiment aims to compare the results obtained by two techniques for segmenting urban paved roads in RGB satellite images (i.e., without thermal band): an unsupervised one, K-Means++ [4], which through the obtained classification allows to segment the region of interest (properly labeled), and a supervised one, SegNet [5]. This study contributes to the development of applications that have common color images as input, which are usually easy and free to access, bringing viability advantages.

Section 2 alludes to scientific works related to the detection of pathways in remote sensing images. Then, in section 3, an overview of the K-Means++ algorithm and the Segnet convolutional neural network is presented. Section 4 details the tools and methodology used in this study. Their results, and their discussion and analysis are shown in section 5. Finally, section 6 concludes the proposed study, pointing out possible improvements and extensions for future work, as well as some of the potential applicability.

2 Related works

The presented situation motivates several works in the scientific literature that propose path detection in remote sensing images. Among them, we can mention Pinho et al [6], that compare different classification methods in urban images with high spatial resolution. Using IKONOS satellite images, the authors concluded that object-oriented classification methods are more suitable than the usual methods for classifying mixed pixels, that is, pixels that belong to more than one class of objects if evaluated by traditional classification algorithms. The use of classes and objects allows the insertion of traditional elements of visual interpretation in the form of descriptors, such as color, shape, size, texture, pattern and context.

In addition to this, Simões [7] uses Artificial Neural Networks (ANN) to segment images through color classification, aiming to maximize the classifier performance by determining a relationship between image attributes and different case studies.

In this line, the work of Venturieri [8] also uses backpropagation algorithms to improve the success rate of using ANNs in land use characterization using color and texture attributes. Despite achieving interesting success rates, the methodology used consumes a lot of processing time and, according to the author, was not suitable for a real-time application.

Nóbrega [1] uses path detection techniques through object-oriented classification in multispectral images with high spatial resolution. Due to the heterogeneity of the present elements in the images, an identification and segregation of the different features into specific classes was carried out, considering the spectral, geometric and contextual characteristics of each one, reaching 64.5% accuracy.

Still in the scope of image classification, Rollet et al [9] combined K-Means centroid initialization procedures with Radial Basis Function ANNs on Landsat TM and MEIS II images (both multispectral), obtaining a more effective classifier than conventional classification methods.

Finally, Doucette et al. [10] also did other notable work in detecting roads in multispectral satellite imagery using an unsupervised method of classification, entitled Self-Organized Road Map (SORM), which was obtained from combination of Self-Organizing Maps and minimum spanning trees.

The most notable practical difficulty in applying the cited works lies in the fact that multispectral images with high spatial resolution are difficult to acquire (usually expensive or restricted), which commonly makes unfeasible the development of applications that requires this kind of input. The proposed study aims to overcome such obstacles by applying and evaluating segmentation methods in RGB images, that is, which do not have thermal band and correspond to those that are usually available in free geoservices and, therefore, easily accessible.

3 Fundamentals

Several works traditionally use K-Means [11] as an image classification method. If, on the one hand, it does not require the hard-working image labeling for being unsupervised, on the other hand, Convolutional Neural Networks (CNNs) do require but are currently considered state-of-the-art in deep learning applied to image processing and pattern recognition. [12]. Such good performance may eventually enable real-time applications to edit geometric features and one that has shown remarkable performance in image segmentation is SegNet [5].

The following sections discuss the main fundamentals of the K-Means++ classifier and the SegNet CNN.

3.1 K-Means++

The images used in this study use the RGB color space, which in turn is based on the Cartesian coordinate system. [13]. Since this is a metric space [14], it is possible to apply distance-based algorithms in the pixels of an RGB image.

The K-Means algorithm [11] partitions a set of n elements into k partitions in which each element belongs to the partition with the closest average. This results in a partitioning of the data space in Voronoi cells.

Voronoi Diagrams are a topological data structure that partitions and discretizes a continuous space in a tessellation[1]. Formally, as explained by Aurenhammer [15], let us consider X a metric space (that is, a set where the distances between any of its elements are defined), non-empty, with a distance function d . Let C_i, with $i \in N$, be an ordered pair (called "centroid") of non-empty subsets in X. The *Voronoi cell* (or *Voronoi region*) V_i, associated with the centroid C is the set of all points in X of which the distance to C_i is not greater than the distance to any other sites C_j, where $j \in N$ is any index other than i. Mathematically,

$$V_i = \{x \in X;\ d(x, C_i) \leq\ d(x, C_j), \forall i, j \in \mathbb{N}, i \neq j\}.$$

The *Voronoi Diagram* $\mathcal{V}$ is the collection of cells V_i, that is,

$$\mathcal{V} = \bigcup_{i=0} V_i.$$

Figure 2 illustrates a set of points partitioned according to the definition of Voronoi cells, which delimit the region of influence of the respective centroid and contain the set of points that are closer to their respective centroid than to any other.

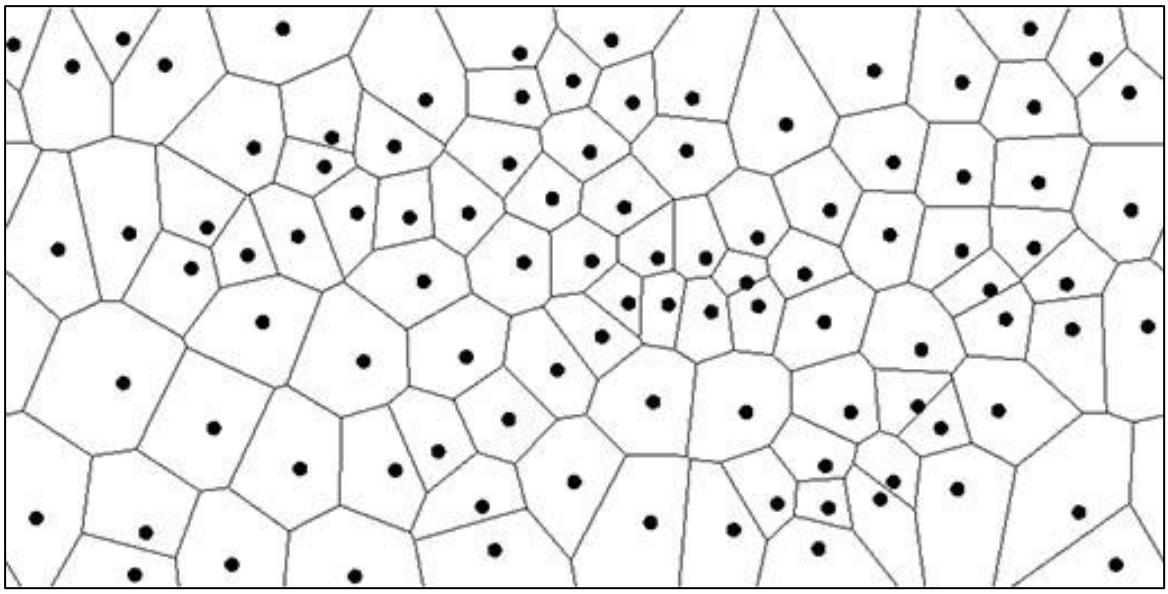


Figure 2. Voronoi diagram, with centroids (black dots) and their respective regions of influence (polygons).

[1] Covering a two-dimensional surface by polygonal basic units, congruent or not, so that there are no spaces or overlaps between them [39].

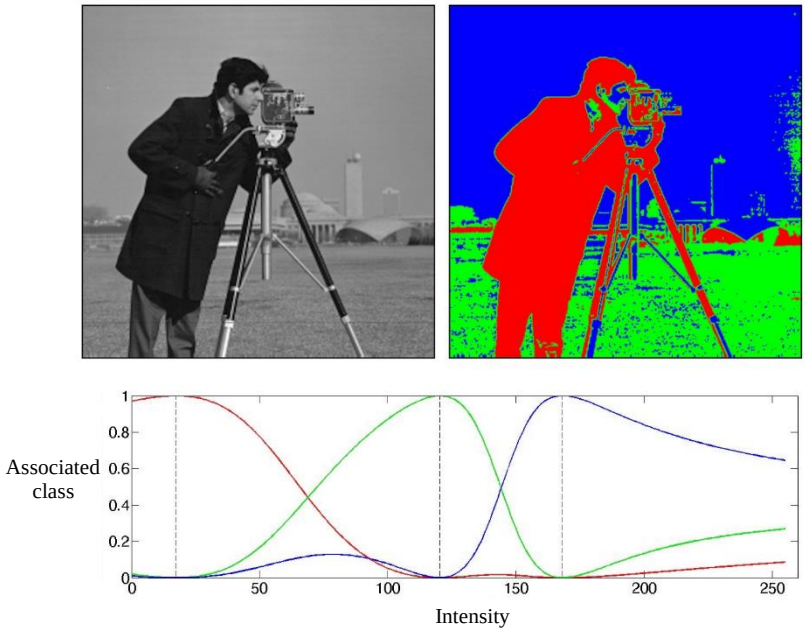


Figure 3. Example of image clustered by K-Means (right), with three classes [16].

The K-Means idealized by MacQueen requires samples as a parameter, which will serve as starting points (that is, centroids) for the clustering process. However, Arthur and Vassilvitskii [4] developed a technique for choosing centroids that considerably improved the speed (often between 100% and 200%) and the accuracy (by at least 20%) of the original K-Means. This improved version is usually called K-Means++ and is the one used in this experiment.

3.2 SegNet

A CNN is a type of ANN designed for minimal computational cost [17] and inspired by the functioning of the human visual cortex [18]. These networks are also known as SIANNs (Shift-Invariant ANNs), because the processing result is independent of the spatial position of the objects in the image [19]. The CNN is formed by a succession of layers as described below.

The *convolutional layer* is responsible for feature extraction from the input volume by convoluting filters over it, generating feature maps, which is a two-dimensional structure that accommodates the responses of this filter in all spatial positions of the input image of the layer. The CNN will learn which filters are activated when they detect a specific feature (edges, blurs, eventually even complete patterns or semantic elements) through *activation functions* [17]. These ones determine, mathematically, whether a neuron should be activated or not depending on the relevance of the information transmitted to it at that moment. On CNNs, the most used activation function is ReLU (*Rectified Linear Unit)*, whose discussion can be found in the work of Googfellow et al [20] and Ramachandran et al [21].

In a CNN architecture is common to periodically insert *pooling layers* between successive convolution layers. Its function is to perform *down sampling*, that is, to progressively reduce the spatial size of the representation in order to reduce the number of parameters to be processed, thus keeping control of overfitting[2] [17]. As in the convolution layer, the pooling layer is obtained by applying a mask over the image.

Additionally, the *up sampling layer* is an idea introduced by Long et al [22] in the development of CNN architectures for image segmentation. In this layer, the image is resized to its original spatial size while enhancing features preserved during down sampling.

[2] *Overfitting*, is a statistical concept that describes the excellent model fitting to a previously known set of data, but ineffective in predicting new results. [42].

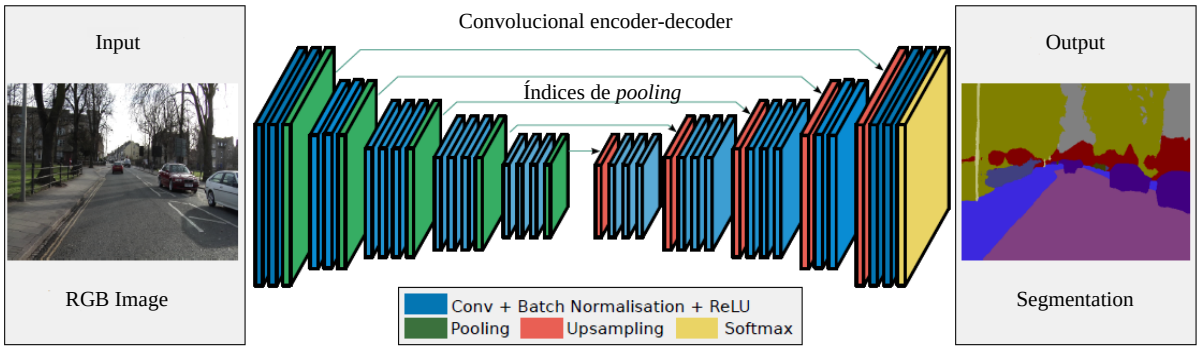


Figure 4. Architecture of a CNN adapted for image segmentation [5].

Although applicable in several areas, such as speech processing and natural language [23], CNNs are widely used in image and video processing. Among the relevant productions, it is possible to cite the one by Ferreira [24], who used CNNs to detect weeds in drone images of soybean crops; or Pereira et al [25], who developed a CNN architecture capable of segmenting brain tumors on MRIs (Magnetic Resonance Images) with enough scores to beat the *Brain Tumor Segmentation Challenge 2013*.

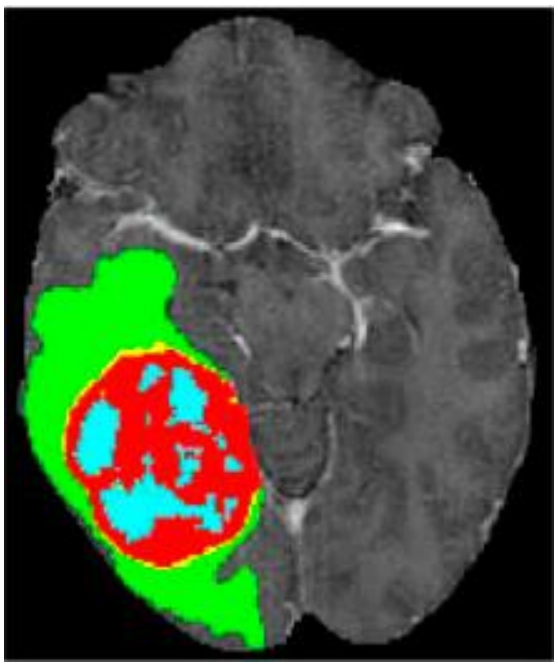


Figure 5. MRI of a segmented brain tumor classified by a CNN, divided by classes: growing tumor (red), stabilized size tumor (yellow), necrosis (blue), edema (green) [25].

Based on the work of Long et al [22], Badrinarayanan et al [5] developed the CNN *SegNet*. Like the CNN that inspired it, SegNet is made up of two parts: the *encoder*, similar to the VGG-16 [26], in which the convolution and pooling operations are done, and the *decoder*, in which the image returns to its original resolution.

The downsampling (or encoding) part of SegNet is similar to a typical convolutional network architecture, but without the dense layers[3]. The basic idea is to add an expansion network later, symmetrical to the contraction one in which pooling operations are replaced by up sampling operators, increasing the resolution of the output until it is equal to the input.

[3] *Fully Connected Layers* or *dense layers* are those in which every neuron is connected with all the others in the next layer, following the same definition as the ANNs of the Multi-Layer Perceptron type [43] [44].

This upsampling (or decoding) step basically consists of two tricks. The first one is the storage of the indexes of pixels selected by *max pooling*. Then, through the transposed convolution, the stored pixels form the resulting image. This provides greater preservation of image characteristics during the encoding and decoding process and substantial memory and storage savings as other CNNs such as U-Net [27], use the entire original image as a feature map. The second trick is regular transposed convolution[4] operations, that occur until the resulting image has the same dimensions as the input image.

This CNN obtained an average IoU of 60.1% when tested in the "CamVid" dataset (competitors did not have this metric extracted) and 90.4% accuracy against 83.8% of the second-best algorithm in this same database. In the "SUN RGB-D" database, the average IoU was 31.84% against 32.08% for the best algorithm, but surpassing all of them in accuracy [5].

4 Methods

4.1 Overview

This study consists of carrying out a series of steps that sequentially produce the proposed experiment, illustrated in Figure 6**.**

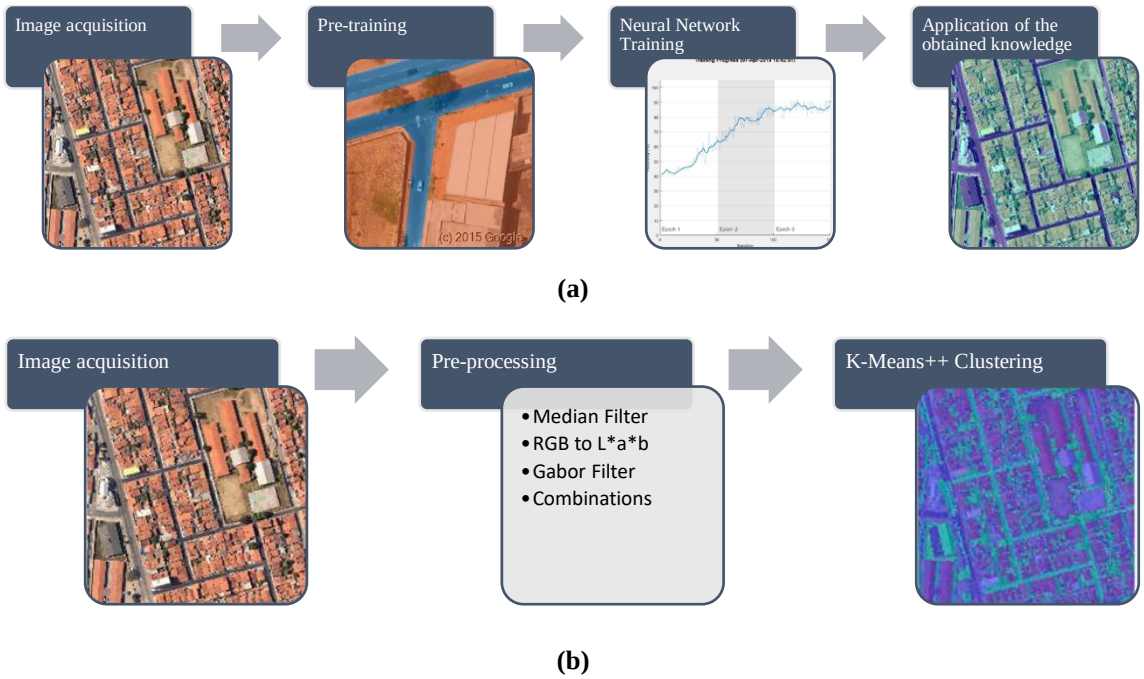


Figure 6. Steps of the proposed work in sequence: a) SegNet; b) K-Means++.

The first flow (Figure 6-a) corresponds to the preparation and use of SegNet. The first step is to obtain the training and testing images for CNN. Then, in the second step, the regions of interest and the training and test image sets are defined. The third stage consists of training SegNet to generate its knowledge bases, enabling them to perform the proposed task. Finally, in the fourth step, by applying CNNs trained in the previous step, the roads segmentation in the provided images is obtained.

Figure 6-b, on the other hand, corresponds to the image processing flow by K-Means++. The first step uses the same images used in the SegNet test. In sequence, color space transformation and image enhancement filters are applied. Finally, in the third step, K-Means++ is applied to segment the roads on the same images used in the first step.

[4] The *transposed deconvolution* is characterized by the use of convolution in upsampling operations, usually done by applying a filter of size $K = 3$ and step $S = 2$ in an image with zero-padding $p = 1$. Each element of the source image is at the center of the filter in an associated step of the convolution [22].

The experiment was implemented on a machine equipped with an Intel Core i7 7700HQ CPU, Nvidia GeForce 1060M GTX GPU, and other configurations that are irrelevant to the present study.

4.2 Image acquisition

In view of the justifications for this study, a base of 600 RGB images with dimensions 256 × 256 pixels and a spatial resolution of 25 cm was used. To facilitate the detection of the roads as much as possible and that there was a minimum amount of mixed pixels, that is, those that belong to several classes in the image.

The images are from urban sectors of São Luís, Maranhão, Brazil (2° 31′ 48″ S, 44° 18′ 10″ O [28]), were captured by Digital Globe and published by Google.

4.3 Database

For each image, a Region of Interest (ROI) is defined, represented by a matrix of integers, with the same dimensions. Each cell (x, y) of this matrix is associated with a pixel of the image and holds a natural number that is associated with a class C, given by

$$C(x,y) = \begin{cases} 0, \text{if the pixel does not belong to any class;} \\ 1, \text{if the pixel corresponds to a road;} \\ 2, \text{if the pixel belongs to a class other than a road.} \end{cases}$$

The ROIs were designed with the help of *Matlab Image Labeler*.

(a) **(b)**

Figure 7. a) The different polluting elements of the image highlighted in rectangles: sandy sediments on the roadway (blue), random afforestation (red), shadows (yellow), concrete tones in asphalt erosion repairs (green) and watermark of the Google (bottom right). b) Image of item a) with ROI street drawn without considering the aforementioned polluting elements.

It should be noted, however, that for practicality, the ROIs were designed without taking into account some elements that pollute the image.

Such elements are a considerable amount of sediments characteristic[5] of the soil of São Luís; random afforestation over considerable portions of some roads; shadows of buildings and trees; varied tones and textures

[5] The soil of Metropolitan Region of São Luís is mainly composed of oxisol and argisol [36], which makes natural the presence of sandy and clayey sediments in the urban landscape in question [37].

on the asphalt caused by precarious paving and subsequent patches, in addition to the watermark generated by Google itself when downloading the images (condition for granting them), and the differences in lighting between quadrants obtained at different times by the satellite.

4.4 Pre-processing methods and K-Means++ application

O K-Means++ is an unsupervised method and has been applied with $k = 2$ clusters in six different ways on the same 100 images that will be used as a test base for SegNet.

Initially, these images were submitted to K-Means++ without any pre-processing (labeled "No"). Then, these images were submitted to different pre-processing, in a non-cumulative way. The first pre-processing consisted of applying the median filter (labeled as "Med") of mask size 15 before classifying them with K-Means++. The objective was to improve the image with noise reduction and mixed pixels, that is, with properties common to different classes. Since K-Means++ works based on distance, the second pre-processing was the transformation of the color space from RGB to CIELAB (labeled as "Lab"), which stores the distance metrics between colors (pixels) in the a and b dimensions, while the L dimension accommodates the luminosity [13]. In sequence, the Gabor filter (labeled as "G") was applied to obtain information about the texture of each pixel.

Finally, under the same conditions, the $Med + Lab$ and $Med + G$ combinations were tried on the test base prior to the execution of K-Means++.

The application of these was done in *Matlab R2020b*.

4.5 SegNet training and application

SegNet learning supervise powered by 500 images with their best ROIs. There were no better image requirements as in the experiment with K-Means ++, that is, as images were submitted to training with No strategy.

Under the same conditions, the remaining 100 to be used by this CNN for application and validation of the generation knowledge base.

SegNet training was done in *Matlab Deep Learning Toolbox*.

4.6 Evaluation metrics

This experiment uses the following quality metrics [29]:

- *Accuracy:* indicates the percentage of correctly identified pixels. This metric indicates how well the method identifies the pixels corresponding to the paved roads. is given by

$$Accuracy = \frac{VP}{VP + FN},$$

  where the denominator of the formula corresponds to the ground truth. The accuracy value considered is the mean between the accuracies obtained for all images in the test set.
- *Jaccard Index:* also called Intersection over Union (IoU). This statistical metric considers and penalizes false positives. It is expressed by

$$IoU = \frac{VP}{VP + FP + FN}.$$

  The considered Jaccard Index is the average between the IoUs obtained for all images in the test set.

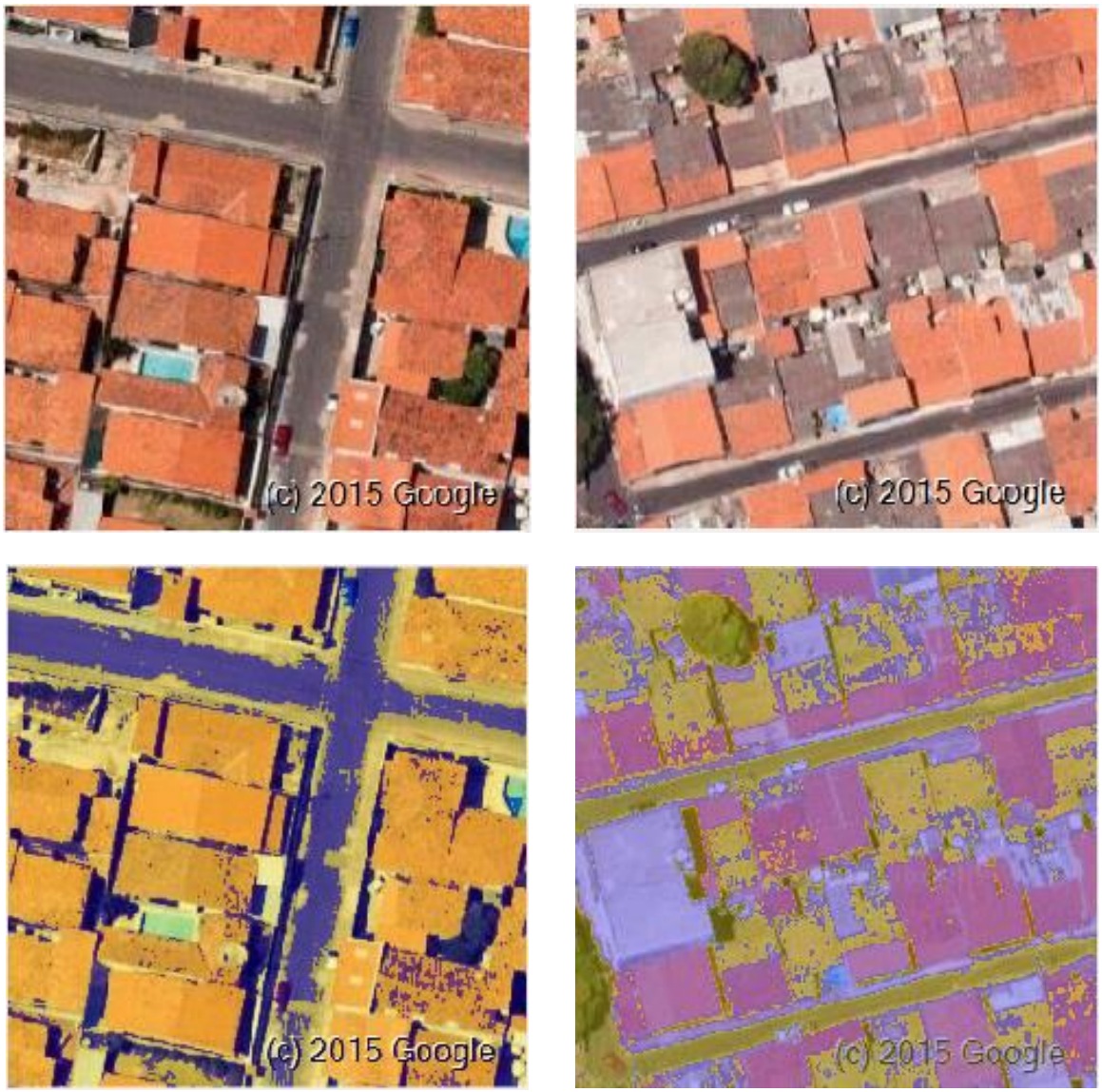
5 Results and discussion

The K-Means++ obtained timid mean accuracy in most cases, ranging from 41.67% ($Med + G$) to 64.19% (Lab). The average IoU ranged between 12.30% ($Med + G$) and 16.16% (Lab).

Except for No and Med applications, in some cases the algorithm barely recognized any or all pixels as belonging to class 1 (worst and best cases, respectively). For the case No the accuracy ranged from 5.84% to 95.90%, while Med ranged from 1.54% to 99.93%.

Taking into account the false positives, the same strategies that obtained the worst accuracies also achieved cases of practically null IoUs. Despite this, these strategies achieved IoUs from 8.80% to 26.67%. On the other hand, for the No and Med strategies, the IoU ranged from 2.15% to 9.89% and 0.47% to 16.69%, respectively.

Such results show that K-Means++ is limited to grouping pixels only by color, which makes that objects with a color close to those of paved roads are indiscriminately included in the same class. Therefore, elements such as fiber cement roofs, shadows, dark treetops and other elements of a darker tone are falsely classified as paved roads.



(a) **(b)**

Figure 8. Examples of urban images and their respective segmentations[6] with K-Means++. The image on the left (a) has few dark colored objects other than asphalt, which provides a coherent classification. On the right (b), on the other hand, it has an incoherent classification and a visible high number of false positives, due to its color close to that of the asphalt.

[6] In Matlab, in K-Means++ threads, the colors associated with each class are random. There is no relationship, therefore, between the color assigned to a class in image segmentation (a) with the same color assigned to a different class in image (b).

Table 1. Comparison between methods: K-Means++ applications (with and without pre-processing) and SegNet.

	Method	Accuracy			IoU		
		Worst case	Average case	Best case	Worst case	Average case	Best case
K-Means++	No	5,84%	47,36%	95,90%	2,15%	4,94%	9,89%
	Med	1,54%	40,01%	99,93%	0,47%	8,95%	16,69%
	Lab	0,01%	98,85%	**99,99%**	0,01%	19,09%	8,80%
	G	0,01%	47,75%	**99,99%**	0,01%	14,49%	12,10%
	$Med + Lab$	0,01%	97,48%	**99,99%**	0,01%	15,33%	25,17%
	$Med + G$	0,01%	31,34%	**99,99%**	0,01%	10,02%	26,67%
CNN	$SegNet$	**73,57%**	**83,26%**	94,73%	**44,96%**	**68,91%**	**90,24%**

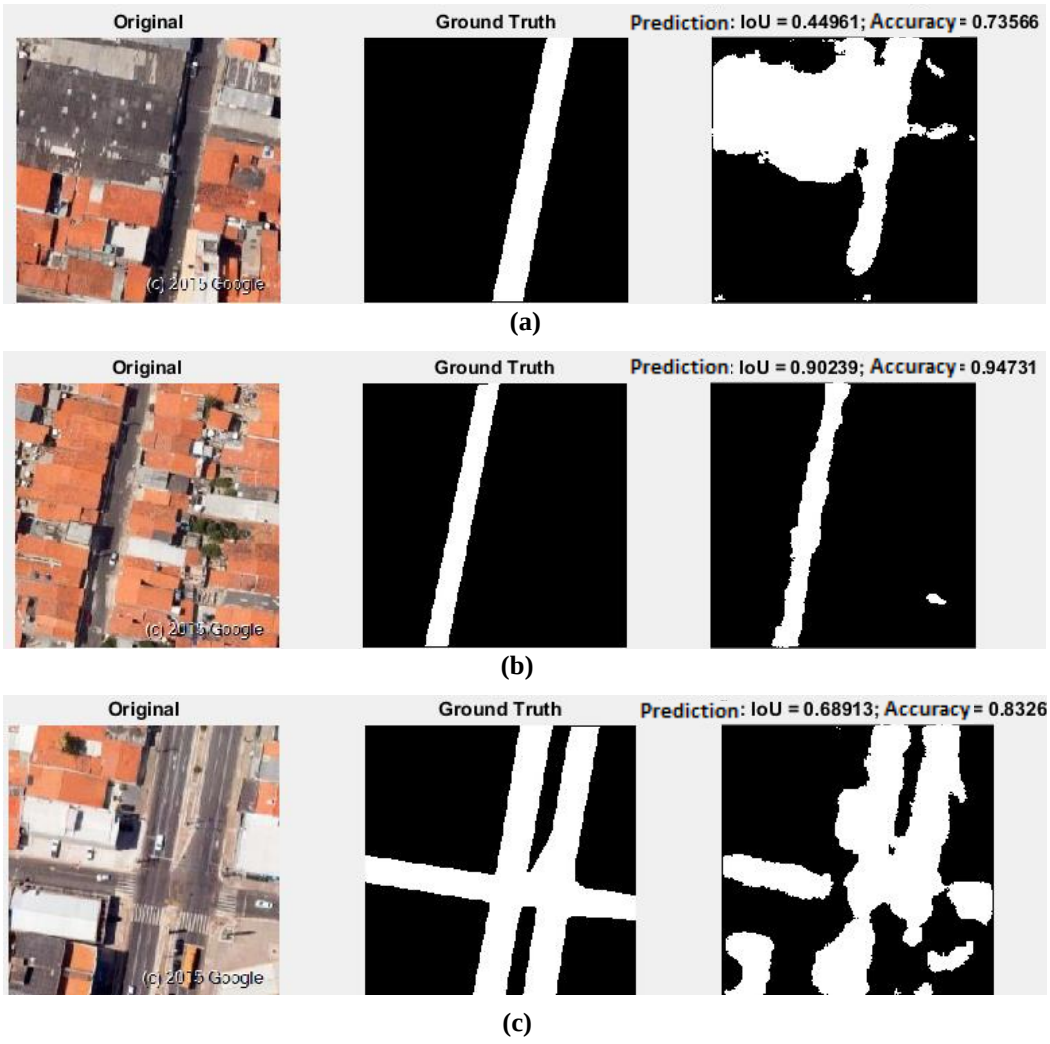


Figure 9. Worst (a), best (b) and average (c) test case for SegNet referring to the “Roads” class.

SegNet, on the other hand, achieved an accuracy of 83.97% and an IoU of 71.93% for road detection. These numbers demonstrate that this CNN was able to differentiate well between the paved roads and the rest of the image in most test cases, largely due to its semantic processing capability.

The worst case achieved an accuracy of 73.57% and an IoU of 44.96%, also showing a considerable amount of false positives, although not comparable to K-Means++. It is possible to observe in the original images corresponding to the worst case and the average case, the presence of fiber cement roofs and concrete slabs, which have a color and texture similar to those of asphalt. In the original image corresponding to the best case, there is a certain uniformity in the asphalt characteristics and, at the same time, great contrast between the characteristics of the other objects that make up the background. There is, therefore, some difficulty in contextualizing the objects and, in this way, differentiating asphalted roads from elements whose color and texture are very similar to asphalt, as can be seen in Figure 9-a.

The best case (Figure 9-b) had an accuracy of 94.73% and an IoU of 90.24%, evidencing satisfactory SegNet performance even with some amount of image polluting elements (in this case, notably shadows and sediment). In the original image corresponding to the best case, there is some uniformity in the features of the asphalt and, at the same time, a great contrast between the features of the other objects that make up the background. This highlights the good performance of the network when there are not so many asphalt-like characteristics, especially in terms of color and texture. Even so, there is a slight amount of false positives in the lower right corner of the image, reinforcing the denotation used in the worst-case analysis.

Similarly, the average case of the sample (Figure 9-c), here, obtained 83.26% of average accuracy and 68.91% of IoU when evaluated by SegNet, and it is possible to note the existence of problems already detected in the worst case.

When considering the general context, that is, taking into account all the classes across the entire base, the K-Means++ achieves an average accuracy of 41.67% to 64.19%, while SegNet achieves an average accuracy of 87.12%.

Including the false positives in the set, K-Means++ achieves an average IoU ranging from 12.30% to 16.16%, depending on the adopted pre-processing strategy, while SegNet achieves an average of 71.93%. This information is shown in Table 2.

Table 2. General comparison between the results of the K-Means++ and SegNet methods.

Method	Average Accuracy	Average IoU
K-Means++	41,67% a 64,19%	12,30% a 16,16%
SegNet	**87,12%**	**71,93%**

These results reveal the unfeasibility of using K-Means++ for the proposed objective, as the numbers indicate a significant amount of false positives and negatives due to the high colorimetric heterogeneity of the images. It is reasonable to state, therefore, that K-Means++, under the conditions presented in this experiment, was not successful for segmenting asphalted roads in RGB images.

On the other hand, despite being very easy to discontinue the segmented region (which can cause problems in certain types of applications, such as those that require skeletonization), SegNet is adequate for the task in question.

6 Conclusions and future work

This experiment aimed to compare and investigate the applicability of two methods for segmenting asphalted roads in RGB images: K-Means++ and SegNet.

K-Means++ achieved poor and unfeasible results for use in a real application due to its inability to semantically analyze the image pixels placement, which makes it difficult to correctly group what is and is not part of an asphalt road. On the other hand, despite the difficulties discussed, the SegNet CNN obtained very superior and encouraging results, affirming the feasibility of using CNNs for segmenting asphalted roads in RGB satellite images, dispensing with the use of multispectral images.

Future work may extend this experiment by improving it in several aspects. For K-Means++, the most relevant ones include verifying the influence of the number of classes, the centroid chosen and studying the possibility of using other attributes besides color by the classifier, so that some semantic information can be aggregated.

As for CNNs, it is desirable and convenient to design more accurate ROIs, preventing regions that do not belong to the roads (especially those mentioned in the section 4.3) from adding confusing or imprecise knowledge to training; and increasing the size of the database, so that CNN has a greater variety of information to train and possibly generate more accurate knowledge. Furthermore, a more in-depth study of the influence of the

configuration of CNNs' hyperparameters on the segmentation result can also promote better results. It is also worth to say that more recent scientific productions, such as those by Guérin et al. [30] and Kanezaki [31], have explored the potential of CNNs in unsupervised methods of learning and classification. This enables an eventual comparison between supervised and unsupervised segmentation methods of objects based on convolutional networks. Unsupervised methods do not require labeling, which makes it possible to use large training bases since it is not necessary to spend a lot of time and effort to designing ROIs.

Furthermore, the study presented has the potential for several applications. Among them, it is worth mentioning the automatic vectorization of the segmented area using skeletonization algorithms, many of them being discussed and compared in the study by Plotze and Bruno [32]. It is also relevant to mention the applicability in studies that require calculating the area corresponding to roads in urban landscapes, as several studies, such as the one by Londe and Mendes [33], relate the quality of urban life with the percentage of built-up area (including paved roads)/green area. Regarding the pathways detection in images similar to the ones in this experiment, the investigation and conception of a specific CNN architecture for such task are also encouraged. Finally, is suggested to evaluate the use of CNNs in the detection of roads of any type, including gravel and dirt roads, which are frequent in rural and some urban areas.

Acknowledgements

Authors thank Professor Artur Bernardo Silva Reis (DESTEC/IFMA, Maranhão, Brazil) for your valuable notes in this manuscript. We thank also Professor Jeff Heaton (Washington University in St. Louis, Missouri, USA), who sparked the seed for the idea of using CNNs in this study through his helpful emails.

References

[1] R. A. d. A. Nóbrega, "Detecção da malha viária na periferia urbana de São Paulo utilizando imagens de alta resolução espacial e classificação orientada a objetos," EPUSP, São Paulo, 2007.

[2] W. T. H. Liu, Aplicações de Sensoriamento Remoto, 2ª ed., Campo Grande, São Paulo: UNIDERP, 2006.

[3] J. M. de Moraes Neto, M. P. Barbosa, M. d. F. Fernandes e M. J. da Silva, "Avaliação da degradação das terras nas regiões oeste e norte da cidade de Campina Grande, PB: um estudo de caso," *Revista Brasileira de Engenharia Agrícola e Ambiental,* vol. 6, nº 1, pp. 180-182, 2002.

[4] D. Arthut e S. Vassilvitskii, "K-means++: The Advantages of Careful Seeding," *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms,* vol. SODA '07, pp. 1027-1035, 2007.

[5] V. Badrinarayanan, A. Kendall e R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, nº 12, p. 2481–2495, Dezembro 2017.

[6] C. M. D. d. Pinho, F. d. F. Feitosa e H. J. H. Kux, "Classificação automática de cobertura do solo urbano em imagem IKONOS: Comparação entre a abordagem pixel-a-pixel e orientada a objetos," *Anais XII Simpósio Brasileiro de Sensoriamento Remoto,* pp. 4217-4224, abril 2005.

[7] A. d. S. Simões, "Segmentação de imagens por classificação de cores: uma abordagem neural," São Paulo, 2000.

[8] A. Venturieri, "Segmentação de imagens e lógica nebulosa para treinamento de uma rede neural artificial na caracterização do uso da terra na região de Tucuruí (PA)," São José dos Campos, 1996.

[9] R. Rollet, G. B. Benie, W. Li e S. Wang, "Image classification algorithm based on the RBF neural network and K-means," *International Journal of Remote Sensing,* vol. 19, nº 15, pp. 3003-3009, 1998.

[10] P. Doucette, P. Agouris, A. Stefanidis e M. Musavi, "Self-organised clustering for road extraction in classified imagery," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 55, nº 5-6, pp. 347-358, 2001.

[11] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* vol. 1: Statistics, pp. 281-297, 1967.

[12] H. Noh, S. Hong e B. Han, "Learning Deconvolution Network for Semantic Segmentation," *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV),* pp. 1520-1528, 07 Dezembro 2015.

[13] R. C. Gonzalez e R. E. Woods, Digital Image Processing, 4ª ed., Upper Saddle River, New Jersey: Pearson, 2018.

[14] S. Lipschutz, Schaum's Outline of Theory and Problems of General Topology, McGraw-Hill, 1965.

[15] F. Aurenhammer, "Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure," *ACM Computing Surveys,* vol. 23, nº 3, pp. 345-405, setembro 1991.

[16] Mathworks, "K-means clustering based image segmentation," Mathworks, [Online]. Available: https://www.mathworks.com/help/images/ref/imsegkmeans.html. [Acesso em 24 dezembro 2020].

[17] J. Patterson e A. Gibson, Deep Learning: a Practitioners's Approach, Sebastopol, Califórnia: O'Reilly, 2017.

[18] M. Eickenberg, A. Gramfort, G. Varoquaux e B. Thirion, "Seeing it all: Convolutional network layers map the function of the human visual system," *Seeing it all: Convolutional network layers map the function of the human visual system,* vol. 152, pp. 184-194, 15 Maio 2017.

[19] W. Zhang, A. Hasegawa, O. Matoba, K. Itoh, Y. Ichioka e K. Doi, "Shift-Invariant Neural Network for Image Processing: Learning and Generalization," *Applications of Artificial Neural Networks III,* vol. 1709, pp. 257-268, 16 Setembro 1992.

[20] I. Goodfellow, Y. Bengio e A. Courville, Deep Learning, MIT Press, 2016.

[21] P. Ramachandran, B. Zoph e Q. V. Le, "Searching for Activation Functions," *CoRR,* vol. abs/1710.05941, 2017.

[22] J. Long, E. Shelhamer e T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *CoRR,* 2015 Março 2015.

[23] R. Collobert e J. Weston, "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning," *Proceedings of the 25th International Conference on Machine Learning,* pp. 160-167, 2008.

[24] J. L. Ferreira, G. L. F. d. Silva, A. B. S. Reis, A. B. Cavalcante, A. C. Silva e A. C. d. Paiva, "Segmentação Automática da Próstata em Imagens de Ressonância Magnética utilizando Redes Neurais Convolucionais e Mapa Probabilístico," *Simpósio Brasileiro de Computação Aplicada à Saúde,* vol. 18, nº 1, 26 Julho 2018.

[25] S. Pereira, A. A. V. Pinto e C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *IEEE Transactions on Medical Imaging,* vol. 35, nº 5, pp. 1240-1251, 4 Março 2016.

[26] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary e A. Agrawal, "Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and Building Materials,* vol. 157, pp. 322-330, 30 Dezembro 2017.

[27] O. Ronneberger, P. Fischer e T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer Assisted Intervention (MICCAI),* pp. 234-241, 18 Maio 2015.

[28] GeoHack, "GeoHack - São Luís (Maranhão)," 2019. [Online]. Available: https://bit.ly/2GmW0F5. [Acesso em 5 março 2019].

[29] G. Csurka, D. Larlus e F. Perronnin, "What is a good evaluation measure for semantic segmentation?," *Proceedings of the British Machine Vision Conference,* p. 32.1–32.11, 2013.

[30] J. Guérin, O. Gibaru, S. Thiery e E. Nyiri, "CNN features are also great at unsupervised classification," *4th International Conference on Artificial Intelligence and Applications,* vol. abs/1707.01700, 2017.

[31] A. Kanezaki, "Unsupervised Image Segmentation by Backpropagation," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing,* pp. 1543-1547, Abril 2018.

[32] R. d. O. Plotze e O. M. Bruno, "Estudo e comparação de algoritmos de esqueletonização para imagens binárias," *IV Congresso Brasileiro de Computação,* pp. 59-64, 2004.

[33] P. R. Londe e P. C. Mendes, "A influência das áreas verdes na qualidade de vida urbana," *Revista Brasileira de Geografia Médica e da Saúde,* vol. 10, nº 18, pp. 264-272, 29 julho 2014.

[34] S. J. Russel e P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed., Upper Saddle River, New Jersey : Prentice Hall, 2009.

[35] C. M. D. d. Pinho, F. C. Silva, L. M. G. Fonseca e A. M. V. Monteiro, "Intra-urban land cover classification from high-resolution images using the C4.5 Algorithm," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. XXXVII, pp. 695-700, julho 2008.

[36] IBGE, "Estado do Maranhão: Pedologia," 2011. [Online]. Available: http://geoftp.ibge.gov.br/informacoes_ambientais/pedologia/mapas/unidades_da_federacao/ma_pedologia.pdf. [Acesso em 9 dez 2018].

[37] Embrapa, Sistema brasileiro de classificação de solos, 2ª ed., Brasília, Distrito Federal: Embrapa, 2006.

[38] A. d. P. Braga, A. P. d. L. F. d. Carvalho e T. B. Ludermir, Redes Neurais Artificiais: Teoria e Aplicações, Rio de Janeiro, Rio de Janeiro: LTC, 2000.

[39] T. Stona, "Tesselação, pavimentação ou mosaico," 2011. [Online]. Available: https://www.ime.usp.br/~thaicia/quasicristais/Tess.html. [Acesso em 1 junho 2017].

[40] J. B. Pacheco Junior, "Uso de Redes Neurais Convolucionais na segmentação de vias urbanas asfaltadas em imagens de satélite RGB: estudo de caso em São Luís-MA," PECS, São Luís, 2019.

[41] Mathworks, "Semantic Segmentation Basics," Mathworks, 7 janeiro 2019. [Online]. Available: https://www.mathworks.com/help/vision/ug/semantic-segmentation-basics.html. [Acesso em 2019 abril 23].

[42] S. Skiena, Calculeted Bets, Cambridge: Cambridge University Press, 2004.

[43] S. Haykin, Neural Networks and Learning Machines, 3ª ed., Upper Saddle River, New Jersey: Pearson Education, 2009.

[44] O. Ludwig Jr. e E. Montgomery, Redes Neurais, Rio de Janeiro, Rio de Janeiro: Ciência Moderna, 2007.