Image-based Semantic Segmentation Network for the Brazilian *Cerrado* based on Public Databases

Daniel C. de Coimbra*, Silas P. W. de Oliveira*,
Dimas A. M. Lemes†, José G. Picolo†,
Guilherme Ribeiro Sales*, Valentino Corso* and Cides S. Bezerra*
*CPQD - Research and Development Center, Campinas, São Paulo, Brazil
Email: {dan.c.coimbra, silaswesleypereiradeoliveira}@gmail.com, {guisales, valenti, cbezerra}@cpqd.com.br
†Pontifical Catholic University, Campinas, São Paulo, Brazil
Email: {dimas.lemes, jose.picolo}@puc-campinas.edu.br

***Abstract*—We have developed fully convolutional networks (FCN) for semantic segmentation of satellite imagery based on five Land Use and Land Cover (LULC) categories: native vegetation, agriculture, pasture, urban region, and waterbody. To this end, we gathered and preprocessed public Brazilian data into an annotated dataset with 26,000 segmented 224 × 224 image patches. We obtained images from the Sino-Brazilian CBERS-04A satellite program and segmentation masks from the Terra Class project (INPE/Embrapa). We performed transfer learning on four backbone models: DeepLabV3+, MobileNet, ResNet-50, and VGG-16. We evaluated their performance with IoU, with the respective scores of 45.96%, 34.40%, 45.58%, and 62.78%. However, our dataset is unbalanced, and a balanced IoU yields scores lower than 20% for all models, indicating specialization on majority classes. Despite these shortcomings, our models' masks have a 16-times higher pixel-density than previously available masks, and taking only images as input, without external data or expert curation.**

***Index Terms*—semantic segmentation, satellite imagery, Brazilian *Cerrado*, fully convolutional, transfer learning**

I. Introduction

Semantic segmentation is a task in image processing which aims to provide a classification for each pixel. Such a classification is called a segmentation mask. [3] It is distinct from instance segmentation, which aims to locate individual items of each class. [20] It is also distinct from image classification, which classifies entire images rather than each of their pixels. [13] Our task of interest is the semantic segmentation of satellite imagery of the Brazilian *Cerrado* into classes pertaining to its surface – what are called Land Use and Land Cover (LULC) classes. [7]

There is great environmental interest in the development of automated surveillance systems for the Earth's immense surface. The decades-long Sino-Brazilian CBERS (China-Brazil Earth Resource Satellite) program provides regular, publicly-available images of Brazil's surface area, spanning millions of square kilometers. [21] Such vast data has great latent utility that can be tapped by being fed to a surveillance neural network to obtain a regularly updated LULC segmentation. In principle, one could monitor wildfires, erosion, deforestation, vegetation patterns, human settlement, irregular mining sites, and other LULC categories of interest. [7] [21] As automated intelligent systems, neural networks are interesting for their high scalability, depending only on the availability of computing power.

Our work has aimed to produce a neural network that can produce a LULC semantic segmentation of the *Cerrado*. We have taken publicly available satellite images from the Sino-Brazilian satellite program CBERS-04A and combined them with also publicly available *Cerrado* LULC segmentation masks produced by the Terra Class team, [8] who is associated with the National Institute for Space Research (INPE) and the Brazilian Agricultural Research Corporation (Embrapa), two major public research institutions in Brazil. Our LULC segmentation involves five classes, namely, native vegetation, agricultural plot, pasture land, urban region, and waterbody.

Besides the Terra Class project, there has been one other concerted effort to produce a complete semantic segmentation mask of the *Cerrado*, namely, one by yet another project affiliated to INPE called Terra Brasilis. [9] Both projects made extensive use of non-visual supplementary data and export human curation to produce their masks. Our networks purportedly improve upon this work pipeline by requiring only a satellite image as an input, doing without external data and human input. Our work has thus intended to develop national technology in a more automated and scalable direction.

Furthermore, we have aimed to produce a segmentation mask of the *Cerrado* with a greater spatial resolution than previously available, which went no further than 32 meters per pixel. Our networks produce masks with the same spatial resolution as the images received as input, which in our case resolve to 8 meters per pixel, providing a mask with a 16-times higher pixel-density. Retraining such networks with 2 meters per pixel images would yield even finer-grained masks.

These improvements are achieved through the application of supervised transfer learning to an FCN architecture based on pretrained models. The data annotation was obtained by aligning *Cerrado* satellite image patches with the Terra Class semantic segmentation mask. As such, our model's accuracy is bounded by the Terra Class mask's own accuracy, and our mean IoU score estimates are also based on our model's agreement to the Terra Class mask. Assessing the latter's accuracy, however, is a daunting task, for it requires an independent standard of judgment, as for instance a sampled evaluation by human experts. We have not found any publication intending to carry out this assessment.

This paper is organized in the following sections. Section II briefly reviews the available work on semantic segmentation, both in general and specialized to agriculture and to the Brazilian *Cerrado*. Section III on methodology discusses the dataset we built by gathering and preprocessing public Brazilian data; the neural network configurations we have experimented with; and the training pipeline we develop on *TensorFlow*. Section IV presents and interprets our models' performances on a quantitative and qualitative basis. Finally, section V summarizes our project's improvements and limitations relative to previous work and points to a future direction of research.

II. Related Works

A number of approaches have been proposed for semantic segmentation. These range from traditional signal processing techniques, such as thresholding, hierarchical histograms, and clustering, up to current neural network models. The latter comprise recurrent, convolutional, graph-based, and transformer network architectures. A comprehensive survey is accessible at [1].

We have experimented with Fully Convolutional Networks (FCNs) for semantic segmentation. Whereas trainable convolutional networks have been available for over three decades [2], it has been less than a decade since FCNs have been formally proposed for semantic segmentation, obtaining around a score 67% mean IoU on a major semantic segmentation dataset (PASCAL VOC). [3] This is not far from the 62.78% mean IoU obtained by our optimal model (VGG-16) on our dataset, although results might differ more starkly if a balanced IoU metric were considered, in which our model scored a mere 18.37%, reflecting our unbalanced dataset (cf. Section III).

As recently as 2021, transformed-based models, known as Visual Transformers (ViTs), have obtained what were then state-of-the-art mean IoU scores on benchmark semantic segmentation datasets: 51.82% on ADE20K, which is considered challenging for its sample variety; a 81.3% score on Cityscapes, focused on urban scenarios; but a mere 59.0% on the aforementioned PASCAL VOC, which was by no means a state-of-the-art performance. [4] These figures serve as a baseline comparison for what can be expected from a semantic segmentation model with current technology.

There has a been a recent surge in machine vision applications to agriculture. A review of techniques based on traditional signal processing and on neural network technology is accessible at [5], together with then state-of-the-art mean IoU scores on a number of agriculture-related semantic segmentation tasks. These can also be useful as an indirect comparison baseline. For instance, a recent FCN model was able to obtain a score of 82.1% mean IoU on the task of distinguish vegetation from non-vegetation. [6] However, it is problematic to compare model performance across different datasets and, moreover, across different tasks. In this case, it is much easier to distinguish the two highly contrasting classes of vegetation and non-vegetation than to perform the finer discrimination between vegetation, agriculture, and pasture which our work intends to do.

Our literature research has revealed a single work purporting to develop a neural network for the semantic segmentation of the *Cerrado*, a Master's dissertation under INPE. [7] Their work was based on a manually labelled image classification dataset comprising 80,000 annotated 256 × 256 image patches distributed across 8 classes. The dataset was developed by previous work in which the author was involved and has been named CerraDataV3. We emphasize that it was an image classification dataset. A semantic classification subset seems to have been created from an 10,000-image balanced subset using automated thresholding techniques. From results presented on its fourth chapter, the dissertation used transfer learning to create two models based respectively on DeepLabV3+ and U-Net (more on these below). Their balanced mean IoU scores for the dataset was, respectively, of only 7.03% and 5.27%, substantially lower than our figure of 18.37% (cf. Table III in Section IV). However, care must be taken to compare performance on different datasets, specially given that our segmentation classes are different.

There have, however, been two projects that produced a segmentation mask for the entire two million square kilometers of the Brazilian *Cerrado*, but without the major input of a neural network. One was the Terra Class project, affiliated with both INPE and Embrapa, which we use for supervised learning. Its segmentation mask comprises 15 LULC classes. [8] The other was the Terra Brasilis project, also affiliated with INPE, with a segmentation mask containing 22 LULC classes. [9] We have decided against using the Terra Brasilis segmentation mask because it is available only in Shapefile (SHP) format, which takes extensive computing resources to transform into raster images suitable for training our networks.

We built our FCN models based on four backbone models, which we list here with a reference to their original research papers: the FCN for semantic segmentation DeepLabV3+ [10] and the CNNs for image classification MobileNet [11], ResNet-50 [12], and VGG-16 [13]. The first of these was available on the *TensorFlow* Model Garden [14] and the latter three could be downloaded from *Keras Applications*. [15] It should also be noted that our FCNs based on CNN image classifiers (the latter three) have an encoding-decoding architecture with skip connections that is very similar to a U-Net [16], which has indeed served as an inspiration to our work.

III. Materials and Methods

A. Dataset

Our work was made possible by the segmentation mask produced by the Terra Class project, affiliated with the aforementioned INPE as well as Embrapa. Their team produced a semantic segmentation mask for the *Cerrado* in 2013 with an updated version in 2020, the latter of which we have employed in our work. Combining information from public geographic databases, image processing techniques, and human expert annotations, the Terra Class team was able to provide a semantic segmentation mask in PNG format for the whole Brazilian *Cerrado*, organized into 15 classes.

To reduce the problem's complexity, we grouped simillar classes among these 15 into only five classes, with class codes ranging from 0 to 4. Our mapping is shown in Table I.

TABLE I
SEMANTIC SEGMENTATION MASK MAPPING

Terra Class	Mapping	Class code
No class	No class	0
Unobserved	No class	0
Miscellaneous	No class	0
Primary native vegetation	Native vegetation	1
Secondary native vegetation	Native vegetation	1
Silviculture	Agriculture	2
Perennial crop	Agriculture	2
Semi-perennial crop	Agriculture	2
Single-cycle temporary crop	Agriculture	2
Multi-cycle temporary crop	Agriculture	2
Recent deforestation	Agriculture	2
Pasture	Pasture	3
Mining	Urban region	4
Urban region	Urban region	4
Non-urban buildings	Urban	4
Waterbody	Waterbody	5

We then juxtaposed the Terra Class segmentation mask to satellite imagery obtained from the CBERS-04A satellite program, which was made possible by georreferenced data.

We hand-selected 10 satellite images of the *Cerrado* from varied periods in 2020. Our criterion was good illumination conditions, low cloud coverage, and as varied a terrain as possible, containing a mixture of vegetation, agriculture, pasture, urban regions, and waterbodies.

Each satellite image is about 90K pixels in width and height. We segmented these into 224 × 224 image patches. We employed multispectral images from WPM cameras with 8 meter-per-pixel spatial resolution. As such, each image patch covers about 3 square kilometers. Each satellite image was segmented into 2,600 image patches, totaling 26,000 images in our dataset. Each image patch was then juxtaposed with its appropriate Terra Class segmentation mask, already converted down to five classes (cf. Table I).

The relative class frequencies for the resulting dataset is exhibited on Table II. As can be seen, the dataset is heavily imbalanced, with native vegetation constituting a majority class. Urban regions and waterbodies ended up severely underrepresented, with the other two classes falling in the middle. We did not implement sampling methods to correct for this imbalance.

TABLE II
DATASET CLASS PREVALENCE

Class code	Class	Prevalence
0	No class	0.7%
1	Native vegetation	81.6%
2	Agriculture	10.3%
3	Pasture	7.0%
4	Urban region	0.1%
5	Waterbody	0.4%

Each item in our dataset consists of an image-mask pair with 224 × 224 pixels. The image is in a three-channel PNG format; the channels selected were the Red (R), Green (G), and Near Infrared (NIR) bands. The NIR spectrum is useful to discriminate between soil, water, and biomass, because in this spectrum their spectral properties are markedly distinct. [17]

The dataset images were randomly separated into train, validation, and test subsets with a 70/15/15 proportion. Since semantic segmentation is a multilabel classification problem – each image contains multiple labels –, it was not trivial to implement stratified sampling to ensure proportional representation of each class in each subset. As such, we did not come to implement such a stratified sampling in our separation. It is important to note that the validation subset was used to select the purported best model architecture, and the test subset was only used to produce the final results presented in this paper.

B. Model architecture

A semantic segmentation neural network inputs an image and outputs a mask with a pixel-wise classification. The image and the mask must have matching dimensions ensured either by the network itself (as we have done) or by downstream post-processing. Among the various possible architectures, we chose to carry out our work with Fully Convolutional Neural Networks (FCNs or FCNNs).

Early in our experiments, it became clear that a custom FCN trained from a random initialization would not reach a satisfactory performance. We therefore decided to work exclusively with pretrained models and apply transfer learning techniques to specialize them to our semantic segmentation problem.

We experimented with two classes of pretrained models. First, we worked with a pretrained FCN already designed to perform semantic segmentation. An FCN is constituted by convolutional layers from end-to-end, with downscaling and upscaling layers in-between. We froze its layers and appended new convolution layers (and scaling layers for dimensionality match) with a random He initialization. The pretrained FCN we employed was DeepLabV3+ [10] downloaded from the *TensorFlow Model Garden*. [14]

Second, we worked with a number of pretrained CNNs for image classification. These are constituted by convolutional layers interceded by downscaling layers, with fully-connected (FC) layers for classification at the end. To embed them into an FCN architecture, their FC layers were removed. To recover the image's dimensionality, we appended a number of convolution layers interceded by upscaling layers, in a process known as deconvolution. [4] As such, the pretrained CNNs act as an encoder segment of an encoder-decoder FCN. We added skip connections between the pretrained model's layers and our new added layers. These are intended to preserve precise spatial information from the input, since the successive downscaling layers act as encoders that may extract useful features, but also act as informational bottlenecks, losing track of precise spatial location information. The pretrained

CNNs we worked with were MobileNet [11], ResNet-50 [12], and VGG-16 [13], downloaded from the *Keras Application* repository. [15]

After constructing a number of networks based on pre-trained models, we looked for the one that produced the best results on our validation set under the limited training time available to our team.

C. Training pipeline

We trained our selected models using a single NVIDIA Tesla T4 GPU in a cloud computing environment using *TensorFlow* 2.15.0 and *Keras* 2.15.0 in Python 3.8.

Our 26,000 images dataset was expanded with standard data augmentation procedures: random rotation, random reflection, and random zooming. Further improvement to our pipeline would add random noise and distortions of the photometric, atmospheric, and radiometric kinds, plus variations in illumination, brightness, and hue, promoting a model that is more robust to real-life data issues. However, from what we could gather, it is not clear that CBERS-04A images have gone through (complete) radiometric and photometric corrections. [18] [19]

To ensure proper class separation, we transformed our ordinal encoded masks, wherein each class is represented by a different natural number, into one-hot encoded masks, which represent classes as binary vectors with a single 1-bit distinguishing each class. This encoding prevents the model from learning falsely that the classes would have an ordinal relationship.

As stated previously, we appended new convolutional layers to the pretrained backbone models. Their setup was a random He initialization, a ReLU activation function, and an ℓ^2 regularization set to 10^{-4}, each followed a batch normalization layer for numerical stability and a dropout layer to prevent overfitting.

The model training was configured with a three-fold cross-validation regime and an exponential decay learning rate schedule with a decay rate of 4% every 3,000 images, coupled to an ADAM optimizer for gradient descent. Our loss function was balanced categorical cross-entropy.

The choice of ℓ^2 regularization value, decay rate, and optimizer were a result of a significant, but by no means exhaustive, hyperparameter search carried out with *Keras Tuner* with 10 training epochs for each variation. We discuss the training results in the succeeding section.

IV. Results and Discussion

After suitable hypertuning, we trained four models based on different pretrained backbone models: DeepLabV3+, MobileNet, ResNet-50, and VGG-16. The models were trained for 55 epochs each, with each epoch consuming about 1 hour of GPU compute time.

It is important to note that none of the models were trained to exhaustion; their performance on the validation set was still on the rise at the training cutoff. Further training was not pursed due to time constraints, as each epoch took about 1 hour of GPU compute time on the setup available to our team. This can be seen in Fig. 1, which presents the training and validation unbalanced mean IoU scores per epoch for VGG-16 – which, as we will see, obtained the best performance in our experiments.

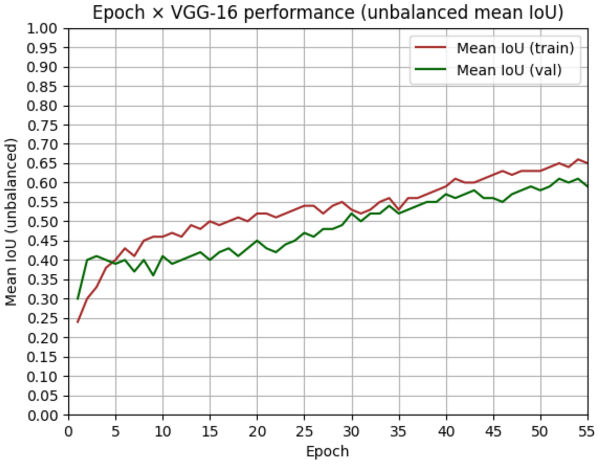

Fig. 1. Unbalanced mean IoU scores for VGG-16 across training epochs. This graph suggests that further training could further improve the model's performance.

As such, the results presented in this section are putative. Our obtained optimal model might well come to underperform given further training to all models. It can, however, be said that our optimal model has been more data efficient, given its superior results in the same number of epochs.

To evaluate the performance of each network, our metric of choice was Balanced Mean Intersection over Union (IoU). The IoU metric is computed for each class and the model is evaluated for its weighed average IoU to account for class imbalance.

Mathematically, the IoU for a single class can be defined as such. Let T and P be the set of pixels attributed to that class by the annotation mask (ground truth) and the predicted mask, respectively. Let $|\cdot|$ designate a set's cardinality. The IoU is defined simply as

$$\text{IoU} = \frac{|T \cap P|}{|T \cup P|} \tag{1}$$

Its value ranges from 0 (empty intersection) to 1 (identity). Intuitively, this measures the percentage of agreement between the two masks, in which false positives (out-class taken as in-class) and false negatives (in-class taken as out-class) count equally.

As previously noted, we selected the best model based on its performance on the validation dataset. The test set was only used to obtain the figures presented in this section, after which no further training was carried out. The results for each model are presented in Table III.

As can be readily seen, the optimal model had a pretrained VGG-16 as its backbone, which is a CNN acting as a feature

TABLE III
TEST SET MODEL PERFORMANCE

Pretrained model	**Mean IoU**	**Balanced Mean IoU**
DeepLabV3+	45.96%	7.59%
MobileNet	35.40%	5.37%
ResNet-50	45.58%	12.17%
VGG-16	62.78%	18.37%

extractor encoding segment. The balanced version of mean IoU is such that each class is weighed in inverse proportion to its dataset prevalence. Note that dataset prevalence was calculated with respect to the whole dataset, and no measure was taken to ensure that the test set formed a representative (stratified) sample of the whole dataset.

It is notable how the introduction of a balanced average, sharply decreases the models' performances. This indicates that every model learned to perform well on the majority classes, but failed to adequately model the minority classes. That our models have learned to specialize on the majority classes can be seen in the confusion matrix for the VGG-16 model in the test set, as can be seen in Table IV.

TABLE IV
TEST SET CONFUSION MATRIX FOR VGG-16

True label	*Predict 1*	*Predict 2*	*Predict 3*	*Predict 4*	*Predict 5*
1	90.1%	4.0%	5.8%	0.0%	0.2%
2	11.1%	80.7%	8.2%	0.0%	0.0%
3	31.0%	12.8%	56.2%	0.0%	0.0%
4	99.8%	0.0%	0.2%	0.0%	0.0%
5	11.7%	0.8%	0.6%	0.0%	86.9%

The confusion matrix displays pixel-wise accuracy scores. Our model was most highly performant on the majority class of native vegetation (label #1). The model's bias towards this class is clear in that it is the most often confounder for other classes, and the minority class of urban region (label #4) is most notable for being almost entirely predicted as native vegetation. Other classes obtained a reasonable accuracy (within around 55% and 85%) and were not often confused for each other, with the moderate exception of a bidirectional confusion between agricultural plots (label #2) and pasture (label #3). A comparison to the class prevalences can be done by referring back to Table I.

Despite this majority class bias, the model as shown itself capable of producing realistic segmentation masks. Fig. 2 exhibits a hand-picked example that showcases the potential of a semantic segmentation model like ours. Many segmentation masks produced by our models are not as realistic, so that the example below should be taken only as an indication of this technology's potential.
The above image pertains to the test set, so that the VGG-16 model had not seen it during training.

Another drawback pertaining to our model's segmentation masks are their narrower class coverage relative to the 15 classes present in the Terra Class masks and the 22 classes present in the Terra Brasilis masks.

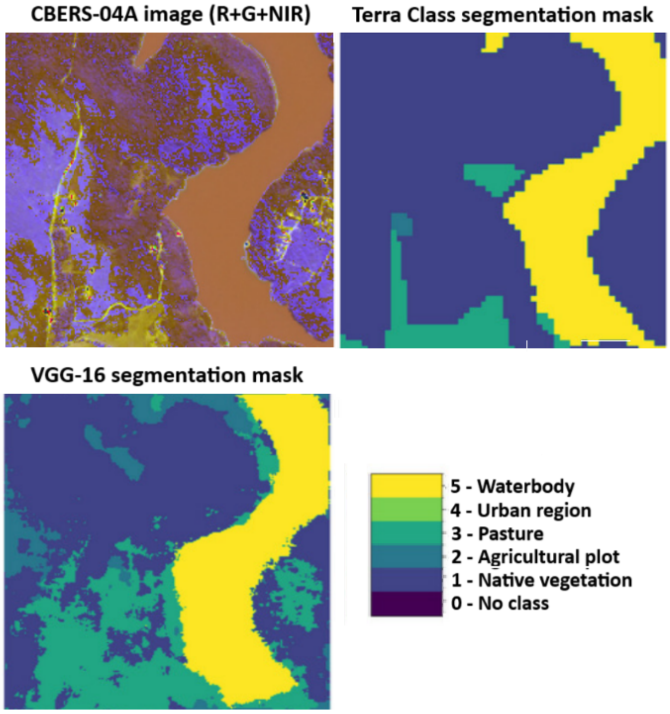

Fig. 2. Example of a figure caption.

Despite the model's significant limitations, it provides a partial improvement upon the Terra Class segmentation mask in that it has a 16-times higher pixel-density. The spatial resolution of the Terra Class mask is about 32 meters per pixel, whereas ours is 8 meters per pixel. Should one train such models with images with a spatial resolution of 2 meters per pixel – which are made publicly available by the CBERS-04A program, albeit in panchromatic form –, the resulting mask would have a 256-times higher pixel-density than the Terra Class mask.

A further improvement provided by our model over the Terra Class project is that the neural network is capable of autonomously producing a segmentation mask based only on image input data, whereas the Terra Class segmentation required manual annotation by a team of experts and the application of external, non-visual geographic data. Plus, although the training dataset has square image patches of uniform size (224 pixels across), a convolutional network can process any rectangular image of a greater size, given enough memory.

As a final remark, it should be highlighted, as foreshadowed in Section I, that our mean IoU metrics are calculated in terms of our model's agreement to the Terra Class segmentation mask. There is a question, however, of the latter's own accuracy with respect to the real world, so the above figures might not translate into real-world performance. Unfortunately, our literature research has returned no independent evaluation of the Terra Class mask's accuracy, as could be performed with a sample evaluation by panel of experts. In addition, given that ours is a supervised learning pipeline, our model's real-world verisimilitude has an upper bound on the Terra Class mask's real-world verisimilitude.

V. Conclusion

We have showcased the feasibility of an image-based semantic segmentation neural network model for the *Cerrado* based on public Brazilian databases with satellite imagery (CBERS-04A) and a semantic segmentation mask (Terra Class).

Our model produces masks that exhibit a 16-times higher pixel-density than previously available masks. Furthermore, our model can produce masks for images of any size and spatial resolution. It can also do so without human intervention and without non-visual data. This constitutes an improvement at least over the Terra Class project's methodology [8], which involved both expert curation and external supplementary data.

Our model's performance, however, is capped by the verisimilitude of the Terra Class segmentation mask used for supervised learning, which has not, as of yet, received an independent assessment.

Our model's balanced mean IoU score of 18.37% also exceeds by a large margin the score of 7.03% obtained by the single previous work in the literature developing a network for the semantic segmentation of the *Cerrado* [7], although the two models cannot be directly because their segmentation tasks involve different LULC classes.

Still, it was not possible to assess the full potential provided either by the public datasets considered in our work or by the Fully Convolutional Network technology, due to time constraints and to limited computation resources available to our team.

In conclusion, our results have indicated considerable latent potential in the application of FCNs to the semantic segmentation of the *Cerrado* using public national data, and serve as groundwork for future research to be carried out with expanded resources.

Acknowledgment

This project was supported by the Ministry of Science, Technology and Innovation, with resources from Law No. 8,248, of October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex and published PDI 03, DOU 01245.023862/2022-14.

References

[1] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, D. Terzopoulos. “Image segmentation using deep learning: a survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 44, n. 7, pp. 3523-42, 2022.

[2] Y. LeCun *et al*. “Handwritten digit recognition with a back-propagation network,” Advances in Neural Information Processing Systems (NeurIPS), 1989.

[3] E. Shelhamer, J. Long, T. Darrell. “Fully convolutional networks for semantic segmentation,” IEEE Conference on Computer Vision and Pattern Recognition (CCVPR), v. 39, n. 4, 2015, pp. 3431–40.

[4] R. Strudel, R. Garcia, I. Laptev, C. Schmid. “Segmenter: transformer for semantic segmentation,” International Conference on Computer Vision (ICCV), 2021.

[5] W. Yang, Y. Yuan, R. Gou, X. Li. “Semantic segmentation of agricultural images: A survey,” Information Processing in Agriculture, v. 10, n. 4, pp. 172-86, 2023.

[6] J. Wang, W. Liu, A. Gou. “Numerical characteristics and spatial distribution of panoramic Street Green View index based on SegNet semantic segmentation in Savannah,” Urban Forestry & Urban Greening, v. 69, 2022.

[7] M. Miranda. “AI4LUC: pixel-based classification of land use and land cover via deep learning and a *Cerrado* image dataset,” Master's dissertation, INPE, 2023.

[8] Terra Class Project. “Mapeamento do uso e cobertura da terra do cerrado,” 2013. Access link.

[9] L. F. Assis *et al*. “TerraBrasilis: A spatial data analytics infrastructure for large-scale thematic mapping,” International Society for Photogrammetry and Remote Sensing (ISPRS) International Journal of Geo-Information, v. 513, n.8, 2019.

[10] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam. “Encoder-decoder with atrous separable convolution for semantic image segmentation,” European Conference on Computer Vision (ECCV), 2018.

[11] A. Howard *et al*. “MobileNets: efficient convolutional neural networks for mobile vision applications,” ArXiv, 2017. Access link.

[12] K. He, X. Zhang, S. Ren, J. Sun. “Deep residual learning for image recognition,” IEEE Conference on Computer Vision and Pattern Recognition (CCVPR), 2015.

[13] Visual Geometry Group (VGG). “Very deep convolutional networks for large-scale image recognition,” International Conference on Learning Representations (ICLR), 2015.

[14] TensorFlow Team. “Model garden - object detection and segmentation.” Access link.

[15] Keras Team. “Keras applications,” Keras 3 API documentation. Acess link.

[16] O. Ronneberger, P. Fischer, T. Brox. “U-Net: convolutional networks for biomedical image segmentation,” International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241, 2015.

[17] European Commission & European Space Agency. “Introduction to remote sensins,” SEOS Project e-Learning Tutorials. Access link.

[18] A. Polidorio, C. Franco, N. Imai, A. Tommaselli, M. Galo. “Correção radiométrica de imagens multiespectrais CBERS e Landsat ETM usando atributos de reflectância de cor,” Anais do XII Simpósio Brasileiro de Sensoriamento Remoto, INPE, pp. 4241-4248, 2018.

[19] T. Akiyama, J. Junior, A. Tommaselli. “Correção geométrica de imagens CBERS-4/PAN com modelos generalizados usando como referência dados do sistema nacional de gestão fundiária,” Anuário de Instituto de Geociências (UFRJ), v. 41, n. 2, 2018.

[20] K. He, G. Gkioxari, P. Dollár, R. Girshick. “Mask R-CNN,” IEEE International Conference on Computer Vision (ICCV), 2017.

[21] CBERS Program. “About CBERS-04A: Uses and Applications,” Brazilian Ministry of Science, Technology, and Innovation. Access link.