Revista Geoaraguaia
ISSN:2236-9716
Barra do Garças – MT
v.15, n.1. Jun-2025

A multi-view evaluation on semantic segmentation supported by deep neural networks for the *Cerrado* savanna

Uma avaliação multivisão sobre segmentação semântica apoiada por redes neurais profundas para o bioma Cerrado

Valdivino Alexandre de Santiago Júnior[1]
Mateus de Souza Miranda[2]
Marcos Lima Rodrigues[3]
Baggio Luiz de Castro e Silva[4]
Thales Sehn Körting[5]

Abstract

While the semantic segmentation task has long been studied by the remote sensing (RS) community, it is a fact that deep neural networks (DNNs) have drawn attention due to the great interest and success of deep learning in several application domains. Even if there are so many studies and experiments using DNNs for RS semantic segmentation, an in-depth multi-view evaluation considering not only different DNNs but also distinct types of images (optical, multispectral) and satellite sensors with diverse spatial resolutions is still missing. This article presents one of such an experimentation where the following options were considered: (1) images from three different

[1] Líder do Laboratório de Inteligência ARtificial para Aplicações AeroEspaciais e Ambientais (LIAREA), Coordenação de Pesquisa Aplicada e Desenvolvimento Tecnológico (COPDT), Instituto Nacional de Pesquisas Espaciais (INPE). Doutor em Computação Aplicada (2011) pelo Programa de Pós-Graduação em Computação Aplicada (CAP) do INPE. Docente do Programa CAP do INPE. E-mail: valdivino.santiago@inpe.br. ORCID: https://orcid.org/0000-0002-4277-021X

[2] Estudante Pesquisador do Laboratório de Inteligência ARtificial para Aplicações AeroEspaciais e Ambientais (LIAREA), Coordenação de Pesquisa Aplicada e Desenvolvimento Tecnológico (COPDT), Instituto Nacional de Pesquisas Espaciais (INPE). Mestre em Computação Aplicada (2023) pelo Programa de Pós-Graduação em Computação Aplicada (CAP) do INPE. Doutorando do Programa CAP do INPE. E-mail: mateus.miranda@inpe.br. ORCID: https://orcid.org/0000-0001-7887-4048

[3] Doutor em Computação Aplicada (2023) pelo Programa de Pós-Graduação em Computação Aplicada (CAP) do Instituto Nacional de Pesquisas Espaciais (INPE). E-mail: marcos.rodrigues@inpe.br. ORCID: https://orcid.org/0000-0002-9199-6928

[4] Mestre em Computação Aplicada (2023) pelo Programa de Pós-Graduação em Computação Aplicada (CAP) do Instituto Nacional de Pesquisas Espaciais (INPE). Doutorando do Programa CAP do INPE. E-mail: baggio.silva@inpe.br. ORCID: https://orcid.org/0009-0009-4226-7196

[5] Pesquisador da Divisão de Observação da Terra e Geoinformática (DIOTG), Coordenação Geral de Ciências da Terra (CGCT), Instituto Nacional de Pesquisas Espaciais (INPE). Pesquisador do Laboratório de Inteligência ARtificial para Aplicações AeroEspaciais e Ambientais (LIAREA). Doutor em Sensoriamento Remoto (2012) pelo Programa de Pós-Graduação em Sensoriamento Remoto (SER) do INPE. Docente do Programa de Pós Graduação em Computação Aplicada (CAP) e do Programa SER do INPE. E-mail: thales.korting@inpe.br. OR CID: https://orcid.org/0000-0002-0876-0501

satellites, i.e., Landsat-8 (30 m of spatial resolution), Sentinel-2 (10 m of spatial resolution), China-Brazil Earth Resources-4A (CBERS-4A; 8 m of spatial resolution); (2) three classical DNNs, i.e., U-Net, DeepLabV3+, PSPNet; and (3) two types of images (optical (RGB) and multispectral). The selected study area was the Brazilian *Cerrado* savanna and the choices of the evaluation focused more on the state-of-the-practice. A thorough investigation was performed and results show that DNNs and spatial resolution of satellite sensors are more relevant than the types of images. This conclusion is interesting because, eventually, researchers may rely on images with less number of channels (optical), decreasing the computational effort during training the DNNs.

Keywords: Deep neural networks; Semantic segmentation; Multi-view evaluation; *Cerrado.*

Resumo

Embora a tarefa de segmentação semântica tem sido estudada há muito tempo pela comunidade de sensoriamento remoto (SR), é fato que as redes neurais profundas (RNPs) têm chamado atenção devido ao grande interesse e sucesso do aprendizado profundo em diversos domínios de aplicação. Mesmo que existam tantos estudos e experimentos utilizando RNPs para segmentação semântica de imagens de SR, ainda falta uma avaliação multivisão aprofundada considerando não apenas diferentes RNPs mas, também, tipos distintos de imagens (ópticas, multiespectrais) e sensores de satélite com diversas resoluções espaciais. Este artigo apresenta uma dessas experimentações onde as seguintes opções foram consideradas: (1) imagens de três satélites diferentes, ou seja, Landsat-8 (30 m de resolução espacial), Sentinel-2 (10 m de resolução espacial), Satélite Sino-Brasileiro de Recursos Terrestres-4A (CBERS-4A; 8 m de resolução espacial); (2) três RNPs clássicas, ou seja, U-Net, DeepLabV3+, PSPNet; e (3) dois tipos de imagens (ópticas (RGB) e multiespectrais). A área de estudo selecionada foi o bioma Cerrado brasileiro e as escolhas da avaliação focaram mais no estado da prática. Uma investigação minuciosa foi realizada e os resultados mostram que RNPs e a resolução espacial dos sensores de satélite são mais relevantes do que os tipos de imagens. Essa conclusão é interessante porque, eventualmente, pesquisadores poderão contar com imagens com menor número de canais (ópticas), diminuindo o esforço computacional durante o treinamento das RNPs.

Palavras-Chave: Redes neurais profundas; Segmentação semântica; Avaliação multivisão; Cerrado.

Introduction

Semantic segmentation of remote sensing (RS) images is a key task that classifies each pixel in an image into a specified category helping to improve visual understanding (SU et al., 2022a). This task has long been addressed by the RS community and even though more recent segmentation tasks, such as instance (GONG et al., 2022) and panoptic (CARVALHO et al., 2022) segmentation, have drawn attention from the RS community, semantic segmentation can still be considered the most popular among them. It has been widely used in many operational RS systems (INPE, 2024a; INPE, 2024b).

The *Cerrado* savanna is the second largest Brazilian biome, after the Amazon biome, corresponding to 23.9% of the national territory, extending over 2,036,448 km^2. It is characterised by typical and strong phytophysiognomies in which woody plants have thick stems, a dark tone, and are twisted but, in other cases, the branches can be angled close to the ground and the tip facing upwards (MIRANDA et al., 2022). Furthermore, the *Cerrado* is home of one of the largest production of agricultural commodities, including both short-cycle and perennial crops, as well as cattle livestock (SANTANA et al., 2020).

The current success of artificial intelligence (AI) is mainly due to deep learning (DL) (MENGHANI, 2023), a subfield of machine learning (ML). Many different deep learning neural networks, or simply deep neural networks (DNNs), models have been proposed to address semantic segmentation of RS images based on convolutional neural networks (CNNs) (CHEN et al., 2022; LIU et al., 2022), attention mechanism-based networks (SU et al., 2022b; LI et al., 2022b) including vision transformer(ViT)-based strategies (LI et al., 2022a), and generative adversarial networks (GANs) (SUN et al., 2021). Moreover, a significant number of datasets have been used for evaluating the DNNs for semantic segmentation such as the Gaofen Image Dataset (GID) (TONG et al., 2020), DeepGlobe (DEMIR et al., 2018), and Wuhan Dense Labelling Dataset (WHDLD) (SHAO et al., 2020).

As it is expected, there are studies which use DNNs, particularly CNNs, to map land use and land cover (LULC) in the *Cerrado* savanna. In (NEVES et al., 2021), the authors presented an adaptation of the architecture of the U-Net network (RONNEBERGER et al., 2015) for processing high spatial resolution images (2 m) to classify the phytophysiognomies of an area of the *Cerrado*. Combinations of spectral bands with vegetation indices were tested in a hierarchical classification. In (MATOSAK et al., 2022), a methodology to detect deforestation in the *Cerrado* via the combination of the long short-term memory (LSTM) and U-Net networks, and using Landsat and Sentinel image time series, was presented. In their approach, the LSTM evaluates the time series in relation to the time axis to create a deforestation probability map, which is spatially analysed by the U-Net model alongside the terrain slope to produce final deforestation maps.

However, the *Cerrado* has a great diversity of landscape arrangements, as presented in the Brazilian TerraClass - *Cerrado* project (INPE, 2024b) which aims to map LULC in anthropised areas identified by the project *Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite* (PRODES—Monitoring Deforestation in the Brazilian Amazon Forest by Satellite) (INPE, 2024a). Thus, there are still gaps in understanding the suitability of DNN models in handling satellite images considering *Cerrado* images. For instance, it is important to take into account image datasets acquired from different satellite/sensors with diverse spatial resolutions. In general, studies use a single dataset (single space resolution) to assess the proposed DNNs and compare to others (SU et al., 2022b; ZHAO et al., 2022; SUN et al., 2021; MARETTO et al., 2021; NEVES et al., 2021; DIAKOGIANNIS et al., 2020; MAXWELL et al., 2020; WANG et al., 2022a; KONG et al., 2021; ANDRADE et al., 2022; WANG et al., 2022b; SUN; ZHENG, 2023). Moreover, several approaches rely on the traditional ISPRS Vaihingen and/or ISPRS Postdam datasets (CHEN et al., 2022; LIU et al., 2022; BAI et al., 2022; SU et al., 2022b; ZHAO et al., 2022; LI et al., 2022a; DIAKOGIANNIS et al., 2020; TASAR et al., 2019). While these datasets have high spatial resolution (indeed, ground sampling distance in these cases) and are relevant, they do not present the features/characteristics of image datasets acquired from the *Cerrado*.

This study presents an in-depth experimentation where the following options were considered: (1) images from three different satellites, i.e., Landsat-8 (30 m of spatial resolution), Sentinel-2 (10 m of spatial resolution), China-Brazil Earth Resources-4A (CBERS-4A; 8 m of spatial resolution); (2) three classical DNNs, i.e., U-Net, DeepLabV3+, PSPNet; and (3) two types of images (optical (RGB) and multispectral). The selected study area was the Brazilian *Cerrado* savanna. The objective of this study is not only to point out which DNN is the most appropriate, but also to understand which of the factors (model itself, spatial resolution, types of images) most influence the performance as a whole for the task of segmenting images from the *Cerrado* savanna.

A broad experimentation process, from different perspectives, is very relevant for the RS community as it provides results and discussion to guide researchers in choosing the most suitable AI solutions for the task of semantic segmentation. Hence, this is the main contribution of this study.

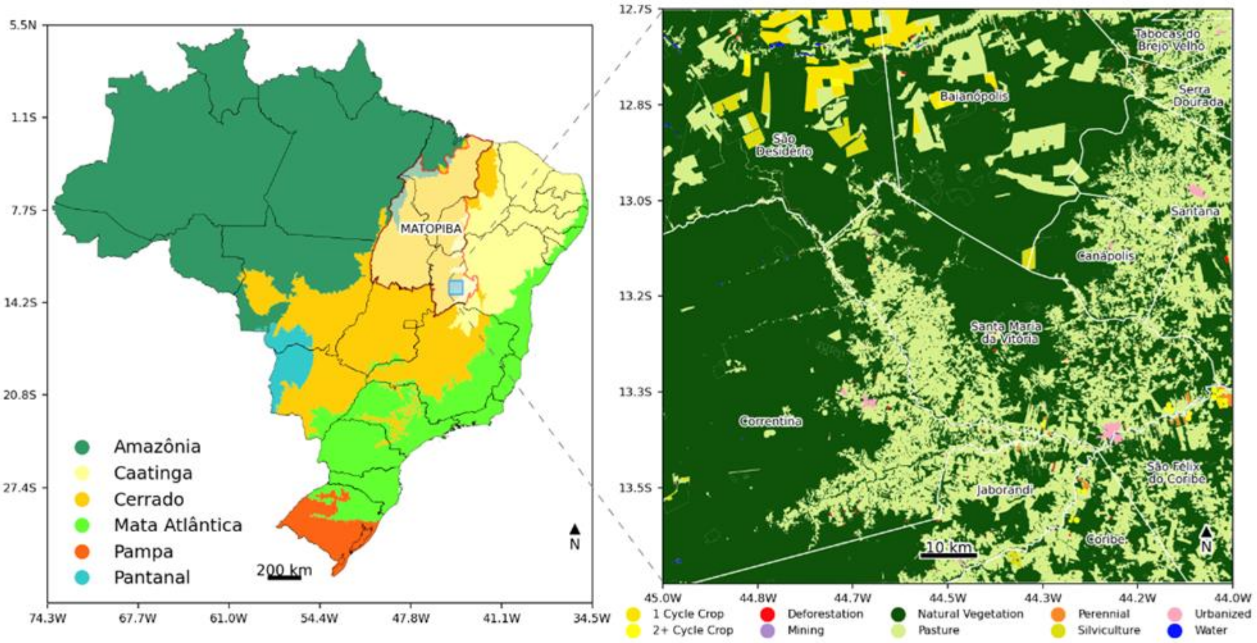
Material and methods

Study area

The *Cerrado* represents the largest savanna in South America, covering about 24% of the Brazilian territory. Some of the most important hydrographic basins of Brazil are located in this biome, such as *Araguaia-Tocantins* and *São Francisco*, as well as the tributaries of the Amazon and Prata rivers. Until the beginning of the 1960s, the *Cerrado* remained relatively isolated from other regions, but the development of new technologies and the creation of government programs to encourage occupation, implemented from the 1970s onwards, transformed it into a new and important Brazilian agricultural frontier (SANO et al., 2019). This transformation changed regional socioeconomic aspects and boosted agricultural productivity, especially in a region composed of the states of *Maranhão*, *Tocantins*, *Piauí*, and *Bahia*, known as MATOPIBA, responsible for much of the Brazilian grain production, with an area of about 73 million hectares (EMBRAPA, 2024), making Brazil one of the world's leading producers of agricultural commodities.

The selected study area is located in MATOPIBA and includes some municipalities in the western region of the *Bahia* state (Figure 1). This region is categorised by heterogeneous LULC, having a predominance of natural vegetation (native) and herbaceous pasture as well as some areas with one or more cycles of temporary crop, silviculture, urbanised areas, deforestation, perennial agriculture, water bodies, and very few mining areas.

Figure 1- Brazilian biomes with the study area located in the MATOPIBA region (left) and zoom in with heterogeneous LULC mapped by the TerraClass - *Cerrado* project for 2020 (right).



Source: Created by the authors (2025)

Classes, satellite information, and datasets

The reference data come from the TerraClass - *Cerrado* project which defines 15 thematic classes (MAURANO et al., 2019). As for this study, some classes were grouped based on similar context resulting in 10 classes (Table 1) with the predominance of Natural Vegetation and Pasture.

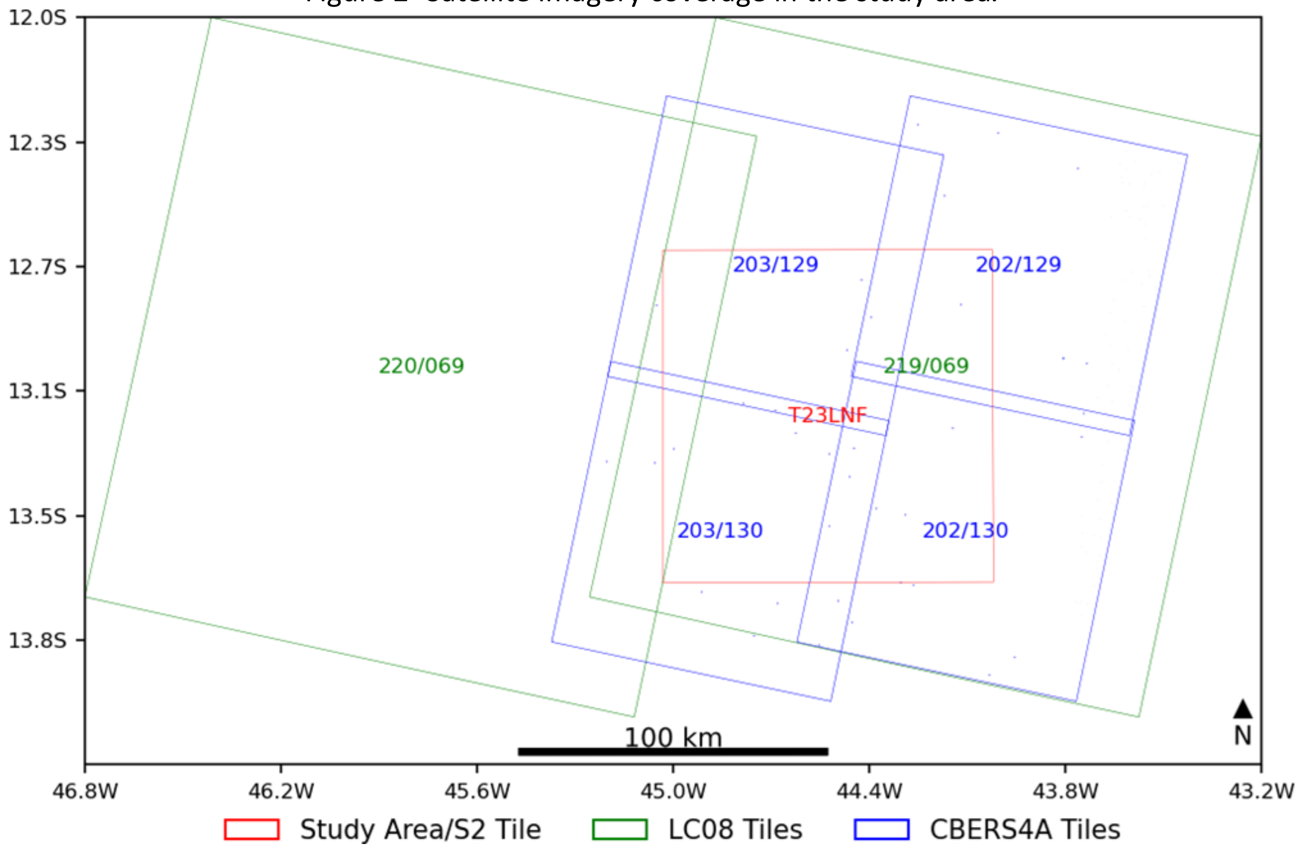
Even if a single area (MATOPIBA) was taken into account, image datasets were created, based on several scenes (Figure 2), with different spatial resolutions by considering the Wide Panchromatic and Multispectral Camera (WPM) of the CBERS-4A satellite (CBERS4A Tiles), the Operational Land Imager (OLI) of the Landsat-8 satellite (LC08 Tiles), and the MultiSpectral Instrument (MSI) of the Sentinel-2 satellite (S2 Tile).

Table 1- Area and percentage coverage of LULC classes in the study area.

LULC Class	**Area (ha)**	**% Area**
Primary and Secondary Natural Vegetation	834,483	69.2
Pasture	334,718	27.8
One Cycle Temporary Crop	17,069	1.4
Silviculture	8,228	0.7
More than One Cycle Temporary Crop	2,798	0.2
Urbanised + Other Built Areas	2,781	0.2
Deforestation	2,648	0.2
Perennial Agriculture	1,252	0.1
Water Bodies	1,242	0.1
Mining	24	< 0.1

Source: Created by the authors (2025)

Figure 2- Satellite imagery coverage in the study area.



Source: Created by the authors (2025)

Several bands, diverse spatial resolutions, and other characteristics from the satellites (Table 2) were considered. The surface reflectance information with atmospheric correction was provided by official producers and available in the public catalogue of Google Earth Engine (GEE). The GEE is a cloud-based platform created to permit processing over long time scales and large spatial extents applied to RS studies (GORELICK et al., 2017). The images were mosaicked and cropped to the study area using Python and the GEE application program interface (API), the dates of images were chosen near to the end of the 2019-2020 crop season because the reference data (TerraClass LULC) corresponds to the year of 2020. Note that bands 11 and 12 (SWIR) of the Sentinel-2 satellite were resampled to 10 m, as provided in the GEE public catalogue.

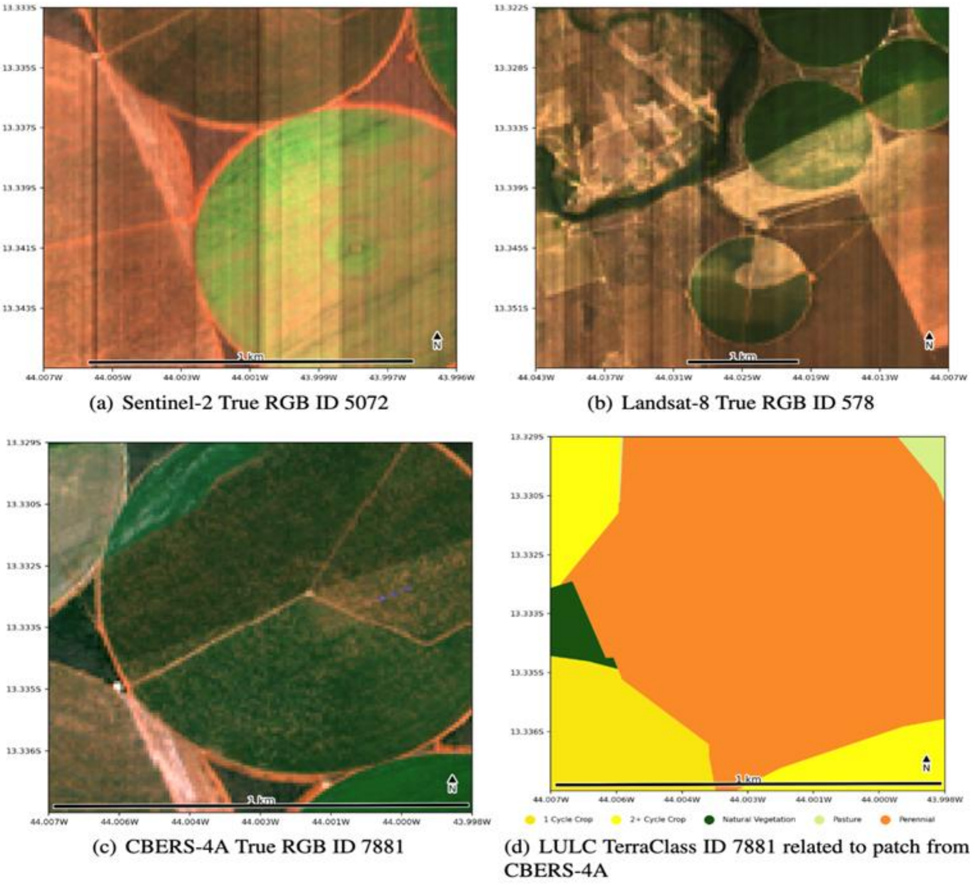
Table 2- Overview of satellite images used. The date follows the format yyyymmdd where y represents year, m means month, and d represents day.

Satellite	Sensor	Path/Row	Date	Bands	Resolution
Landsat-8	OLI	219/69, 220/69	20200826, 20200905	B2:B7	30 m
CBERS-4A	WPM	202/129, 202/130, 203/129, 203/130	20200703, 20200803, 20200829	B1:B4	8 m
Sentinel-2	MSI	T23LNF	20200827, 20200829	B2:B4, B8, B11, B12	10 m

Source: Created by the authors (2025)

Based on the year 2020, patches of 128 x 128 pixels without overlapping and preserving geospatial information were created. The size of the patches was chosen according to a typical resolution (dimension) that is used for input to the selected models aiming to capture enough spatial and semantic context and facilitate the training process. A total of 841, 7,397 and 11,664 patches (Figure 3) were gathered from Landsat-8, Sentinel-2, and CBERS-4A, respectively.

Figure 3- Image patches and mask for the created datasets. Each dataset is made up of RGB/multispectral bands patches and LULC mapped (mask) by the TerraClass - *Cerrado* project.



(a) Sentinel-2 True RGB ID 5072

(b) Landsat-8 True RGB ID 578

(c) CBERS-4A True RGB ID 7881

(d) LULC TerraClass ID 7881 related to patch from CBERS-4A

Source: Created by the authors (2025)

Deep neural networks

Three traditional DNNs were selected. U-Net (RONNEBERGER et al., 2015) is still likely the most popular DNN for image semantic segmentation, where so many studies, including for the RS community, rely on this model or some variations of it. U-Net consists of a contracting path to capture context and a symmetric expanding path that enables precise localisation (U-shape). In other words, it is an encoder-decoder architecture. Such a DNN makes strong use of data augmentation techniques (MUMUNI; MUMUNI,2022) such as shift and rotation invariance which are suitable for microscopical images, since the model participated in a challenge for segmentation of neuronal structures in electron microscopic stacks.

DeepLabV3+ (CHEN et al., 2018) considers two types of neural networks that use spatial pyramid pooling (SPP) module (HE et al., 2014) or encoder-decoder structure (RONNEBERGER et al., 2015) for semantic segmentation, where the former one captures rich contextual information by pooling features at different resolutions while the latter one is able to obtain sharp object boundaries. DeepLabV3+ extends a previous model, DeepLabV3, by employing an encoder-decoder structure. The encoder component of the system processes multi-scale contextual information by performing atrous convolution at various scales, while the straightforward and effective decoder module enhances the segmentation outcomes by refining them along the boundaries of objects.

PSPNet (ZHAO et al., 2017) was proposed for scene parsing and it embeds difficult scenery context features in a fully convolutional network (FCN)-based (LONG et al., 2015) pixel prediction framework. In addition to traditional dilated FCN for pixel prediction, PSPNet extends the pixel-level feature to the specially designed global pyramid pooling one. They also proposed an optimisation strategy with deeply supervised loss.

Metrics and experiment runs

Two metrics were considered in this study: weighted mean Intersection over Union (mIoU) and F1-score. As the name implies, the weighted mIoU is a weighted average of the IoU of each class where the number of pixels of each class in the image are taken into account. The F1-score is the classical metric which can be interpreted as a harmonic mean of the precision and recall.

In all cases, models were trained from scratch for three times, the best model was selected for the inference phase, and the metric used (weighted mIoU, F1-score) is the average of the three executions.

Moreover, the goal is not only to know how many times a particular DNN or type of image or satellite (spatial resolution) is better than other. The point is to realise how much better a solution is. Thus, based on the average of the metrics (weighted mIoU, F1-score), the Improvement metric (SANTIAGO JÚNIOR, 2023), *I%*, was calculated as follows (Equation 1), where *B* and *W* are the best and worst value of the average of the metric, respectively, comparing each pair of solutions (model, type of image, satellite):

$$I\% = \frac{(B-W)\times 100}{W} \quad (1)$$

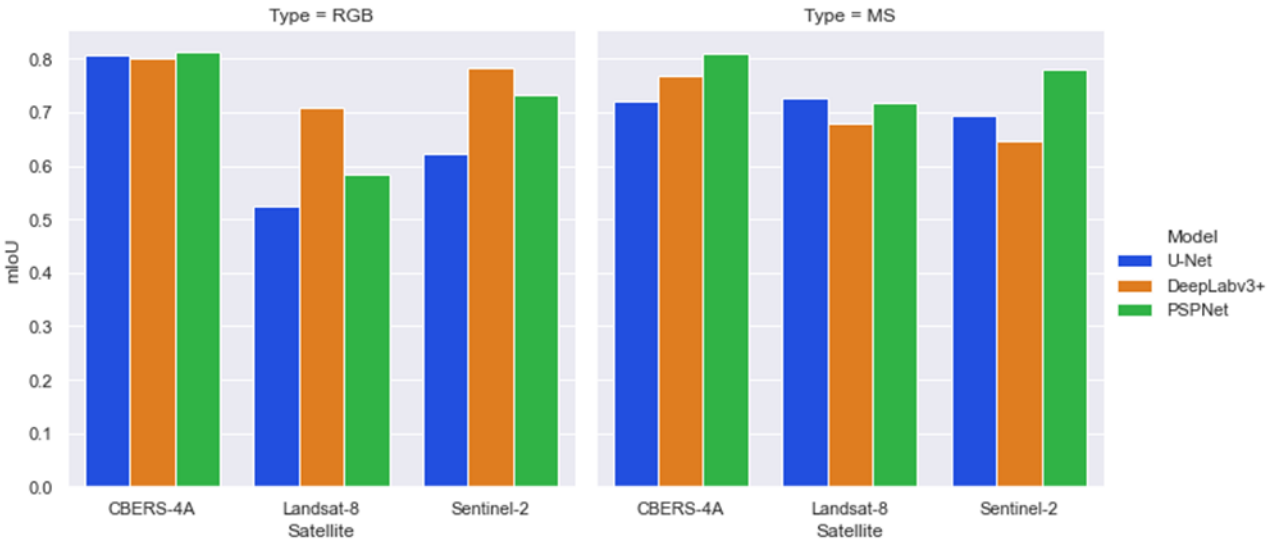
Results and discussion

In this section, the results of the in-depth evaluation are presented. Each perspective is considered independently, as detailed below.

Models

As for the models, their performances were analysed considering each combination/configuration of type of image and satellite regarding the weighted mIoU (Figure 4). Note that the best average of the weighted mIoU (0.812) was obtained by PSPNet with optical (RGB) images and the CBERS-4A satellite. On the other hand, the worst (0.5235) was due to U-Net with optical (RGB) images and the Landsat-8 satellite.

Figure 4- Model perspective: Average of the weighted mIoU. Caption: MS = Multispectral.



Source: Created by the authors (2025)

In the table containing the results considering the *I%* metric with respect to the weighted mIoU (Table 3), the column **Best** shows the number of times the model in the column **Model** was the best, **Max *I%*** means the maximum obtained *I%* by the model in the column **Model**, the column **Max Mod** represents the other compared model where the greatest improvement was seen considering the model in the column **Model**, columns **Max Type** and **Max Sat** mean the type of image and satellite, respectively, where such greatest improvement occurred. Finally, the column ***Avg_I%*** is the average *I%*. Hence, for instance, PSPNet was the best in three configurations with a maximum improvement of 20.552953% compared to DeepLabV3+ with multispectral images and the Sentinel-2 satellite. Its average improvement, *Avg_I%*, was 8.856601%.

Table 3- Model perspective: Results considering *I%* with respect to the weighted mIoU. Caption: MS = Multispectral; Lan = Landsat-8; Sen = Sentinel-2.

Model	Best	Max *I%*	Max Mod	Max Type	Max Sat	Avg_*I%*
U-Net	1	7.192336	DeepLabV3+	MS	Lan	4.335669
DeepLabV3+	2	35.300860	U-Net	RGB	Lan	22.455835
PSPNet	3	20.552953	DeepLabV3+	MS	Sen	8.856601

Source: Created by the authors (2025)

Thus, PSPNet and DeepLabV3+ presented the best performances overall. However, even if PSPNet obtained the highest number of best performances (three), DeepLabV3+ obtained the highest maximum performance (35.300860%) and the highest Avg_*I%*, almost three times higher than the average of PSPNet. On the other hand, in a direct comparison, PSPNet was four times better than DeepLabV3+ while DeepLabV3+ was better twice.

Regarding the F1-score (Figure 5 and Table 4), basically similar results as in the case of the weighted mIoU were seen. PSPNet and DeepLabV3+ presented the best performances overall, the highest maximum performance (34.063052%) and the highest Avg_*I%* were due to DeepLabV3+. However, PSPNet was four times better than DeepLabV3+ while DeepLabV3+ was better twice. Note that even the respective values in columns **Max Mod**, **Max Type**, and **Max Sat** in the F1-score table (Table 4) are the same as in the weighted mIoU table (Table 3). Just as in the weighted mIoU, the best average of the F1-score (0.8592) was obtained by PSPNet with optical (RGB) images and the

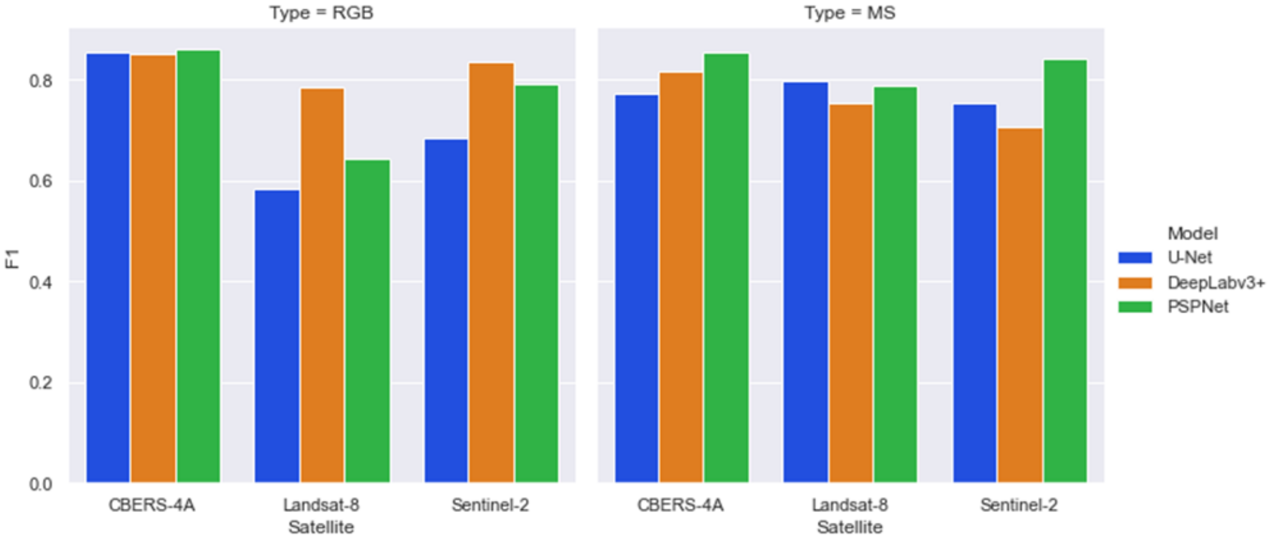
CBERS-4A satellite. The worst (0.5847) was due to U-Net with optical (RGB) images and the Landsat-8 satellite.

Table 4- Model perspective: Results considering *I%* with respect to the F1-score. Caption: MS = Multispectral; Lan = Landsat-8; Sen = Sentinel-2.

Model	**Best**	**Max *I%***	**Max Mod**	**Max Type**	**Max Sat**	**Avg_*I%***
U-Net	1	6.105991	DeepLabV3+	MS	Lan	3.760263
DeepLabV3+	2	34.063052	U-Net	RGB	Lan	20.888136
PSPNet	3	18.926314	DeepLabV3+	MS	Sen	7.924483

Source: Created by the authors (2025)

Figure 5- Model perspective: Average of the F1-score. Caption: MS = Multispectral.



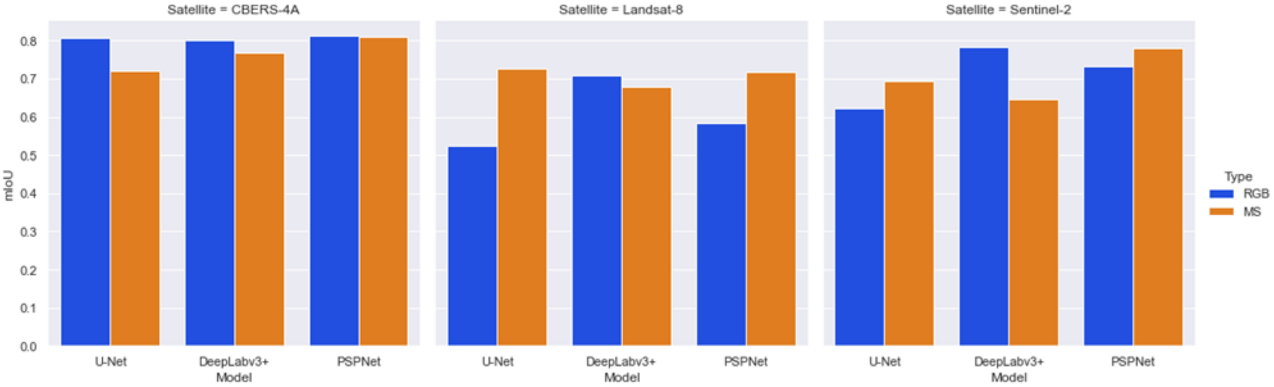
Source: Created by the authors (2025)

Types of image

As for the types of image, their performances were analysed considering each configuration of model and satellite regarding the weighted mIoU (Figure 6 and Table 5).

Note that in the table that represents the performance of the types of image (Table 5), the column **Max Mod** now represents the model where the maximum improvement occurred and the column **Max Type** indeed means the compared type of image where the greatest improvement was seen. Multispectral images obtained the highest maximum improvement (38.930277%) and this happened with the U-Net model and the Landsat-8 satellite. In this case, the type of image that multispectral images were compared to can only be RGB. Multispectral images also obtained the highest Avg_*I%*.

Figure 6- Types of image perspective: Average of the weighted mIoU. Caption: MS = Multispectral.



Source: Created by the authors (2025)

Table 5- Types of image perspective: Results considering *I%* with respect to the weighted mIoU. Caption: MS = Multispectral; Lan = Landsat-8; Sen = Sentinel-2.

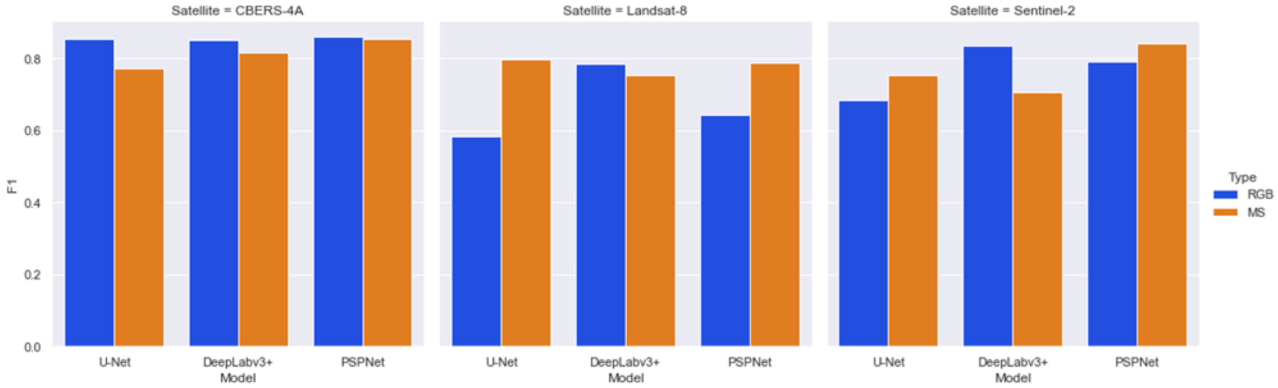
Type	**Best**	**Max *I%***	**Max Mod**	**Max Type**	**Max Sat**	**Avg_*I%***
RGB	5	21.181074	DeepLabV3+	MS	Sen	8.404982
MS	4	38.930277	U-Net	RGB	Lan	19.939194

Source: Created by the authors (2025)

Although the multispectral images performed better than the optical (RGB) ones, it was expected that such performance to be much higher. Note that the optical images had a greater number of better results (five) than the multispectral ones (four), although optical images presented a lower Avg_*I%*.

The results related to the F1-score (Figure 7 and Table 6) show again an agreement with the outcomes presented with the weighted mIoU. Multispectral images obtained the highest maximum improvement (36.514452%) and also Avg_*I%*, but their performance is not as high as it was supposed to be, since optical images obtained the highest number of wins (five).

Figure 7- Types of image perspective: Average of the F1-score. Caption: MS = Multispectral.



Source: Created by the authors (2025)

Table 6- Types of image perspective: Results considering *I%* with respect to the F1-score. Caption: MS = Multispectral; Lan = Landsat-8; Sen = Sentinel-2.

Type	**Best**	**Max *I%***	**Max Mod**	**Max Type**	**Max Sat**	**Avg_*I%***
RGB	5	18.402679	DeepLabV3+	MS	Sen	7.596140
MS	4	36.514452	U-Net	RGB	Lan	18.734271

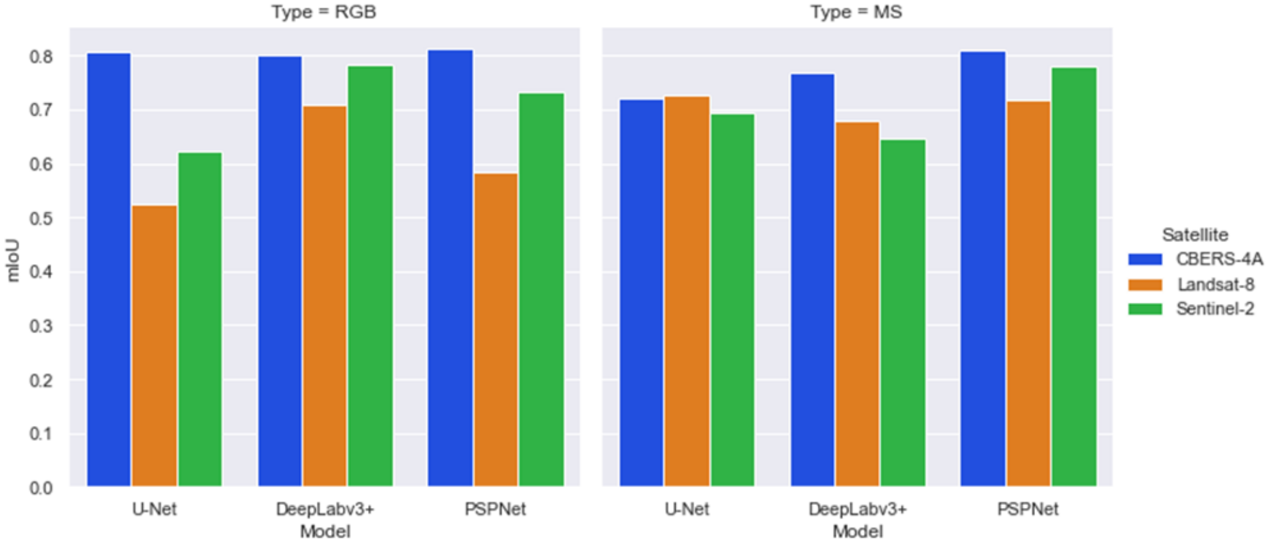
Source: Created by the authors (2025)

Satellites

As for the satellites (spatial resolution) perspective, their performances were evaluated for each configuration of type of image and model with respect to the weighted mIoU (Figure 8 and Table 7). As in the last subsection, in the table that represents the performance of the satellites (Table 7), the column **Max Mod** now represents the model where the maximum improvement occurred and the column **Max Sat** indeed means the compared satellite where the greatest improvement was seen.

It is clear here that CBERS-4A (the satellite/sensor with the best spatial resolution: 8 m) was quite superior than the others wining five out of six configurations with a maximum improvement of 54.00191% compared to Landsat-8 with the U-Net and optical images.

Figure 8- Satellite perspective: Average of the weighted mIoU. Caption: MS = Multispectral.



Source: Created by the authors (2025)

Table 7- Satellite perspective: Results considering *I%* with respect to the weighted mIoU. Caption: MS = Multispectral; Lan = Landsat-8; Sen = Sentinel-2.

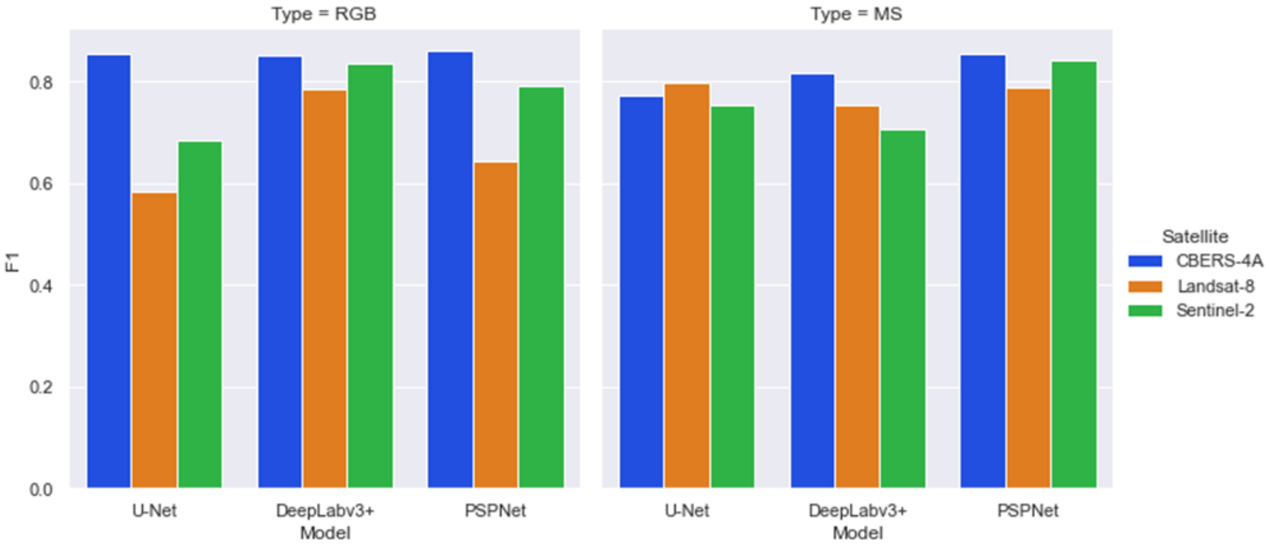
Satellite	**Best**	**Max *I%***	**Max Mod**	**Max Type**	**Max Sat**	**Avg_*I%***
CBERS-4A	5	54.00191	U-Net	RGB	Lan	19.798529
Landsat-8	1	4.803305	U-Net	MS	Sen	2.864209
Sentinel-2	0	-	-	-	-	-

Source: Created by the authors (2025)

Surprisingly, Landsat-8 (the satellite/sensor with the poorest spatial resolution: 30 m) was the best in one situation even surpassing CBERS-4A. In this case, the maximum *I%* was 4.803305% compared to Sentinel-2 with U-Net and multispectral images. However, in a direct comparison between Landsat-8 and CBERS-4A, only in this single situation Landsat-8 was better. Moreover, when comparing Landsat-8 to Sentinel-2 (the one with the intermediate spatial resolution: 10 m), only in two configurations Landsat-8 was better than Sentinel-2. Thus, in the other four configurations, Sentinel-2 obtained better results.

Once more there is an agreement between the weighted mIoU and the F1-score metrics (Figure 9 and Table 8). As for the F1-score, CBERS-4A was quite superior than the other satellites with worse spatial resolutions, and Landsat-8 was the best in one configuration. But, when comparing directly Sentinel-2 to Landsat-8, Sentinel-2 obtained four wins and Landsat-8 obtained only two wins.

Figure 9- Satellite perspective: Average of the F1-score. Caption: MS = Multispectral.



Source: Created by the authors (2025)

Table 8- Satellite perspective: Results considering *I%* with respect to the F1-score. Caption: MS = Multispectral; Lan = Landsat-8; Sen = Sentinel-2.

Satellite	Best	Max *I%*	Max Mod	Max Type	Max Sat	Avg_*I%*
CBERS-4A	5	46.092013	U-Net	RGB	Lan	15.743536
Landsat-8	1	5.824642	U-Net	MS	Sen	4.606980
Sentinel-2	0	-	-	-	-	-

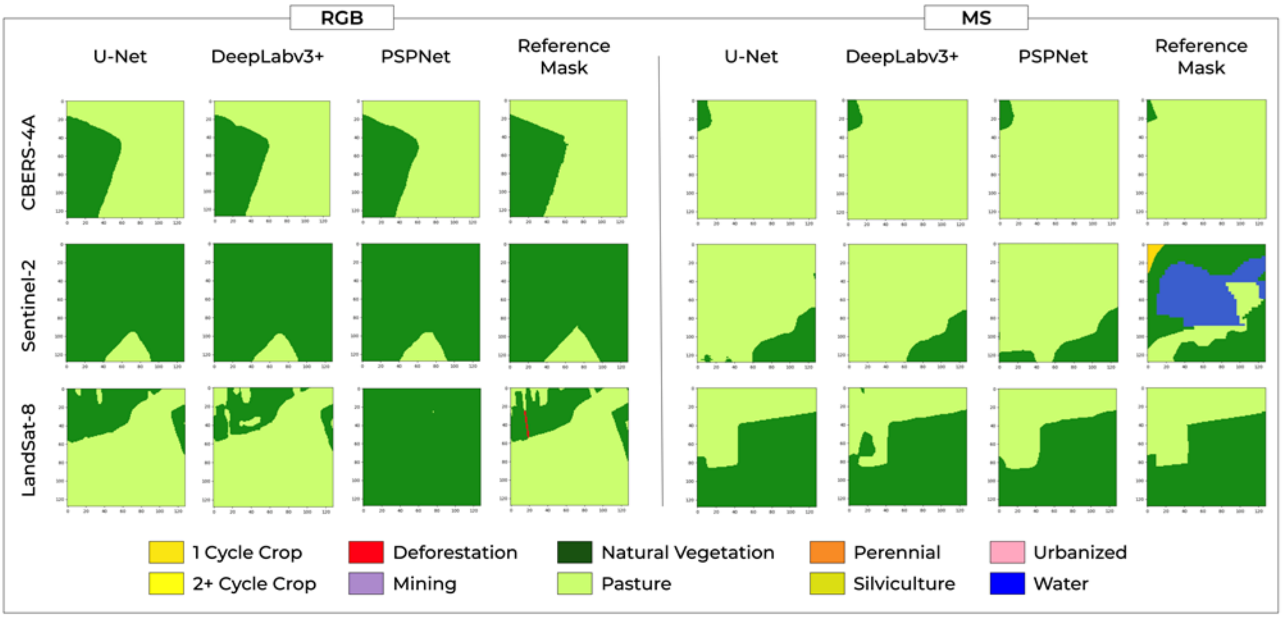
Source: Created by the authors (2025)

Visual analysis

In this study, the models were trained with unbalanced datasets in terms of the number of pixels from each class. Consequently, there is a good chance that the DNNs presented overfitting concerning the Natural Vegetation and Pasture classes (Figure 10). This effect is illustrated in the multispectral (MS) Sentinel-2's reference mask whose models under-performed assigning only Natural Vegetation and Pasture, even for Water Bodies and One Cycle Temporary Crop classes.

Another case is illustrated in the third line of RGB outcomes, where a small-scale Deforestation red polygon is not detected by none of the models. DeepLabV3+ presented better details results whereas PSPNet could not segment the elements adequately as expected, considering these samples.

Figure 10- Some predictions from the DNNs.



Source: Created by the authors (2025)

Very good outputs were achieved by the models with the CBERS-4A dataset, corroborating what was previously stated. The models can predict more visual information given the higher spatial resolution images from CBERS-4A, even when it comes to small regions, as visualised in the MS example. Sentinel-2 and Landsat-8 datasets, however, were not as good as the CBERS-4A dataset.

In general, consistent with the previously presented quantitative results, the models exhibited superior performance with the CBERS-4A dataset, particularly PSPNet and DeepLabV3+. Hence, higher spatial resolution yields more discernible visual information. Concerns exist regarding the appropriate learning of semantic information for certain categories due to the diversity and number of samples available for training.

Conclusions

Since the current significant number of deep learning techniques/methods/DNNs that have been used by the RS community, it is relevant to perform independent evaluations to recommend the most adequate solutions for researchers and professionals. Thus, in this article, the focus was on the task of semantic segmentation of images from the Brazilian *Cerrado* savanna where an evaluation from different perspectives was carried out. This assessment took into account three different points of view, i.e., models/DNNs, types of image, satellites (spatial resolution), trying to identify which perspective is the most relevant.

By considering images from three different satellites/sensors with diverse medium spatial resolutions (Landsat-8, Sentinel-2, CBERS-4A), three traditional DNNs (U-Net, DeepLabV3+, PSPNet), and two types of images (optical (RGB) and multispectral), the conclusion is that DNNs and spatial resolution of satellite sensors are more relevant than the types of images. In other words, based on the experiment that was conducted, using multispectral images may eventually not worth the value, and researchers and professionals may rely on optical (RGB) images to perform semantic segmentation of images from a biome like the *Cerrado* savanna. Even if more experimentation is required, this conclusion is interesting because practitioners may consider images with less number of channels (optical), decreasing the computational effort during training the DNNs.

Regarding the models, DeepLabV3+ and PSPNet were superior than the very popular U-Net. And the performance of CBERS-4A (the satellite with the best spatial resolution) was very much higher than the other ones.

As it is known, pure CNNs can only capture local information, due to the localisation of the convolutional operation. As for future directions, more DNNs will be considered, particularly the self-attention-based approaches which can capture global information, and a comparison to the traditional CNN models will be carried out. Moreover, other types of images, such as hyperspectral ones, as well as images from other high resolution satellite/sensors will be taken into account and new experiments will be conducted.

Acknowledgements

This research was developed within the project *Classificação de imagens e dados via redes neurais profundas para múltiplos domínios* (Image and data classification via Deep neural networks for multiple domainS—**IDeepS**). The **IDeepS** (available online: https://github.com/vsantjr/IDeepS (accessed on March 31, 2025)) project is supported by the *Laboratório Nacional de Computação Científica* (LNCC—National Laboratory for Scientific Computing, MCTI, Brazil) via resources of the SDumont supercomputer. This study was also supported in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES—Coordination for the Improvement of Higher Education Personnel, process number: 88887.603929/2021-00), by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) via grant number 302205/2023-3, and by the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP) via grant number 2023/09118-6.

References

ANDRADE, R. B.; MOTA, G. L. A.; COSTA, G. A. O. P. Deforestation detection in the amazon using deeplabv3+ semantic segmentation model variants. **Remote Sensing**, v. 14, n. 19, 2022. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/14/19/4694. Accessed on: March 11, 2024.

BAI, H.; CHENG, J.; HUANG, X.; LIU, S.; DENG, C. Hcanet: A hierarchical context aggregation network for semantic segmentation of high-resolution remote sensing images. **IEEE Geoscience and Remote Sensing Letters**, v. 19, p. 1–5, 2022.

CARVALHO, O. L. F.; CARVALHO JÚNIOR, O. A.; ROSA E SILVA, C.; ALBUQUERQUE, A. O.; SANTANA, N. C.; BORGES, D. L.; GOMES, R. A. T.; GUIMARÃES, R. F. Panoptic segmentation meets remote sensing. **Remote Sensing,** v. 14, n. 4, 2022. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/14/4/965. Accessed on: March 11, 2024.

CHEN, L.; DOU, X.; PENG, J.; LI, W.; SUN, B.; LI, H. Efcnet: Ensemble full convolutional network for semantic segmentation of high-resolution remote sensing images. **IEEE Geoscience and Remote Sensing Letters**, v. 19, p. 1–5, 2022.

CHEN, L.-C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: FERRARI, V.; HEBERT, M.; SMINCHISESCU, C.; WEISS, Y. (Ed.). **Computer Vision – ECCV 2018**. Cham: Springer International Publishing, 2018. p. 833–851. ISBN 978-3-030-01234-2.

DEMIR, I.; KOPERSKI, K.; LINDENBAUM, D.; PANG, G.; HUANG, J.; BASU, S.; HUGHES, F.; TUIA, D.; RASKAR, R. Deepglobe 2018: A challenge to parse the earth through satellite images. In: **The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops**. 2018.

DIAKOGIANNIS, F. I.; WALDNER, F.; CACCETTA, P.; WU, C. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. **ISPRS Journal of Photogrammetry and Remote Sensing,** v. 162, p. 94–114, 2020. ISSN 0924-2716. Available at: https://www.sciencedirect.com/science/article/pii/S0924271620300149. Accessed on: March 11, 2024.

EMBRAPA. **Matopiba**. 2024. Available at: https://www.embrapa.br/tema-matopiba. Acessed on: March 1, 2024.

GONG, Y.; ZHANG, F.; JIA, X.; MAO, Z.; HUANG, X.; LI, D. Instance segmentation in very high resolution remote sensing imagery based on hard-to-segment instance learning and boundary shape analysis. **Remote Sensing,** v. 14, n. 1, 2022. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/14/1/23. Accessed on: March 11, 2024.

GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google earth engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, Elsevier, 2017. Available at: doi.org/10.1016/j.rse.2017.06.031. Accessed on: March 11, 2024.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: FLEET, D.; PAJDLA, T.; SCHIELE, B.; TUYTELAARS, T. (Ed.). **Computer Vision – ECCV 2014**. Cham: Springer International Publishing, 2014. p. 346–361. ISBN 978-3-319-10578-9.

INPE. **Monitoramento do Desmatamento da Floresta Amazônica Brasileira por Satélite (PRODES)**. 2024. Available at: http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes. Acessed on: December 4, 2024.

INPE. **TerraClass**. 2024. Available at: https://www.terraclass.gov.br/. Acessed on: December 4, 2024.

KONG, Y.; LIU, Y.; YAN, B.; LEUNG, H.; PENG, X. A novel deeplabv3+ network for sar imagery semantic segmentation based on the potential energy loss function of gibbs distribution. **Remote Sensing**, v. 13, n. 3, 2021. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/13/3/454. Accessed on: March 11, 2024.

LI, W.; GAO, H.; SU, Y.; MOMANYI, B. M. Unsupervised domain adaptation for remote sensing semantic segmentation with transformer. **Remote Sensing**, v. 14, n. 19, 2022. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/14/19/4942. Accessed on: March 11, 2024.

LI, Z.; ZHANG, Z.; CHEN, D.; ZHANG, L.; ZHU, L.; WANG, Q.; CHEN, S.; PENG, X. Hcrb-msan: Horizontally connected residual blocks-based multiscale attention network for semantic segmentation of buildings in hsr remote sensing images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 15, p. 5534–5544, 2022.

LIU, Z.; LI, J.; SONG, R.; WU, C.; LIU, W.; LI, Z.; LI, Y. Edge guided context aggregation network for semantic segmentation of remote sensing imagery. **Remote Sensing**, v. 14, n. 6, 2022. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/14/6/1353. Accessed on: March 11, 2024.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE Computer Society, 2015. p. 3431–3440. ISSN 1063-6919.

MARETTO, R. V.; FONSECA, L. M. G.; JACOBS, N.; KÖRTING, T. S.; BENDINI, H. N.; PARENTE, L. L. Spatio-temporal deep learning approach to map deforestation in amazon rainforest. **IEEE Geoscience and Remote Sensing Letters**, v. 18, n. 5, p. 771–775, 2021.

MATOSAK, B. M.; FONSECA, L. M. G.; TAQUARY, E. C.; MARETTO, R. V.; BENDINI, H. N.; ADAMI, M. Mapping deforestation in cerrado based on hybrid deep learning architecture and medium spatial resolution satellite time series. **Remote Sensing**, v. 14, n. 1, 2022. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/14/1/209. Accessed on: March 11, 2024.

MAURANO, L. E. P.; ALMEIDA, C. A.; MEIRA, M. B. Monitoramento do Desmatamento do Cerrado Brasileiro por Satélite – PRODES Cerrado. In: **Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto.** 2019. p. 191–194.

MAXWELL, A. E.; BESTER, M. S.; GUILLEN, L. A.; RAMEZAN, C. A.; CARPINELLO, D. J.; FAN, Y.; HARTLEY, F. M.; MAYNARD, S. M.; PYRON, J. L. Semantic segmentation deep learning for extracting surface mine extents from historic topographic maps. **Remote Sensing**, v. 12, n. 24, 2020. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/12/24/4145. Accessed on: March 11, 2024.

MENGHANI, G. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 55, n. 12, mar 2023. ISSN 0360-0300. Available at: https://doi.org/10.1145/3578938. Accessed on: March 11, 2024.

MIRANDA, M. S.; SILVA, L. F. A.; SANTOS, S. F.; SANTIAGO JÚNIOR, V. A.; KÖRTING, T. S.; ALMEIDA, J. A high-spatial resolution dataset and few-shot deep learning benchmark for image classification. In: **2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. 2022. v. 1, p. 19–24.

MUMUNI, A.; MUMUNI, F. Data augmentation: A comprehensive survey of modern approaches. **Array,** v. 16, p. 100258, 2022. ISSN 2590-0056. Available at: https://www.sciencedirect.com/science/article/pii/S2590005622000911. Accessed on: March 11, 2024.

NEVES, A. K.; KÖRTING, T. S.; FONSECA, L. M.; SOARES, A. R.; GIROLAMO-NETO, C. D.; HEIPKE, C. Hierarchical mapping of brazilian savanna (cerrado) physiognomies based on deep learning. **Journal of Applied Remote Sensing**, SPIE, v. 15, n. 4, p. 044504, 2021.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). **Medical Image**

Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4.

SANO, E. E.; ROSA, R.; SCARAMUZZA, C. A. M.; ADAMI, M.; BOLFE, E. L.; COUTINHO, A. C.; ESQUERDO, J. C. D. M.; MAURANO, L. E. P.; NARVAES, I. S.; OLIVEIRA FILHO, F. J. B.; SILVA, E. B.; VICTORIA, D. C.; FERREIRA, L. G.; BRITO, J. L. S.; BAYMA, A. P.; OLIVEIRA, G. H.; BAYMA-SILVA, G. Land use dynamics in the Brazilian Cerrado in the period from 2002 to 2013. **Pesquisa Agropecuária Brasileira,** v. 54, 2019. ISSN 1678-3921. Available at: https://www.scielo.br/j/pab/a/GQYcBYK993bQksGszxwkcsf/?lang=en#. Accessed on: March 11, 2024.

SANTANA, C. A. M.; CAMPOS, S. K.; MARRA, R.; ARAGÃO, A. A. Cerrado: pilar da agricultura brasileira. In: BOLFE E. L.; SANO, E. E.; CAMPOS, S. K. (Ed.). **Dinâmica agrícola no cerrado: análises e projeções**. Brasília: Embrapa, 2020. p. 39–58.

SANTIAGO JÚNIOR, V. A. Evaluating deep learning techniques for blind image super-resolution within a high-scale multi-domain perspective. **AI**, v. 4, n. 3, p. 598–619, 2023. Available at: https://www.mdpi.com/2673-2688/4/3/32. Accessed on: March 11, 2024.

SHAO, Z.; ZHOU, W.; DENG, X.; ZHANG, M.; CHENG, Q. Multilabel remote sensing image retrieval based on fully convolutional network. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 318–328, 2020.

SU, Y.; CHENG, J.; BAI, H.; LIU, H.; HE, C. Semantic segmentation of very-high-resolution remote sensing images via deep multi-feature learning. **Remote Sensing**, v. 14, n. 3, 2022. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/14/3/533. Accessed on: March 11, 2024.

SU, Y.-C.; LIU, T.-J.; LIUY, K.-H. Multi-scale wavelet frequency channel attention for remote sensing image segmentation. In: **2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)**. 2022. p. 1–5.

SUN, S.; MU, L.; WANG, L.; LIU, P.; LIU, X.; ZHANG, Y. Semantic segmentation for buildings of large intra-class variation in remote sensing images with o-gan. **Remote Sensing**, v. 13, n. 3, 2021. ISSN 2072-4292. Available at: https://www.mdpi.com/2072-4292/13/3/475. Accessed on: March 11, 2024.

SUN, Y.; ZHENG, W. Hrnet- and pspnet-based multiband semantic segmentation of remote sensing images. **Neural Comput & Applic,** v. 35, p. 8667–8675, 2023. Available at: https://link.springer.com/article/10.1007/s00521-022-07737-w. Accessed on: March 11, 2024.

TASAR, O.; TARABALKA, Y.; ALLIEZ, P. Incremental learning for semantic segmentation of large-scale remote sensing data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 12, n. 9, p. 3524–3537, 2019.

TONG, X.-Y.; XIA, G.-S.; LU, Q.; SHEN, H.; LI, S.; YOU, S.; ZHANG, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. **Remote Sensing of Environment,**

v. 237, p. 111322, 2020. ISSN 0034-4257. Available at: https://www.sciencedirect.com/science/article/pii/S0034425719303414. Accessed on: March 11, 2024.

WANG, K.; FAN, X.; WANG, Q. Fpb-unet++: Semantic segmentation for remote sensing images of reservoir area via improved unet++ with fpn. In: **Proceedings of the 2022 6th International Conference on Innovation in Artificial Intelligence**. New York, NY, USA: Association for Computing Machinery, 2022. (ICIAI '22), p. 100–104. ISBN 9781450395502.

WANG, M.; DU, H.; XU, S.; SURNAME, G. N. Remote sensing image segmentation of ground objects based on improved deeplabv3+. In: **2022 IEEE International Conference on Industrial Technology (ICIT)**. 2022. p. 1–6.

ZHAO, H.; SHI, J.; QI, X.; WANG, X.; JIA, J. Pyramid scene parsing network. In: **2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE Computer Society, 2017. p. 6230–6239. ISSN 1063-6919.

ZHAO, Q.; LIU, J.; LI, Y.; ZHANG, H. Semantic segmentation with attention mechanism for remote sensing images. **IEEE Transactions on Geoscience and Remote Sensing**, v. 60, p. 1–13, 2022.