Cloud Segmentation in Multispectral Images from the Sentinel-2 Satellite: A Comparative Study of Deep Learning Approaches

Jean A. C. Dias[1], Williane G. S. Pereira[1], Waldemiro J. A. G. Negreiros[1], Pedro H. do V. Guimarães[1], Alan B. S. Corrêa[1], Leonardo de O. Tamasauskas[1], Marivan S. Gomes[3], Danilo Souza[4], William Yoshida[4], Danilo Valente[4] Daniela Rocha[4], Fernando Augusto[1], Gabriel B. Costa[2], Marcos C. da R. Seruffo[1,2]

[1] Laboratório de Pesquisa Operacional (LPO)
Universidade Federal do Pará (UFPA), Belém-PA

[2]Programa de Pós-Graduação em Estudos Antrópicos na Amazônia
Universidade Federal do Pará (UFPA), Castanhal-PA

[3]Núcleo de Robótica e Automação
Escola Superior de Tecnologia
Universidade do Estado do Amazonas (UEA), Manaus-AM

[4]CARBONEXT, São Paulo-SP

{jean.dias,alan.correa}@itec.ufpa.br, Waldemiro.negreiros@ifpa.edu.br
{williane.pereira, leonardo.tamasauskas}@icen.ufpa.br, seruffo@ufpa.br
{danilo.souza, william.yoshida, danilo.valente}@carbonext.com.br
pedro.guimaraes@castanhal.ufpa.br, msgomes@uea.edu.br,
daniela.rocha@carbonext.com.br, fernando.augusto.mlr@gmail.com,
gabrielbritocosta@gmail.com

Abstract. *To enhance the processing of multispectral images under adverse conditions, such as the presence of clouds, this study compared the neural networks U-Net, DeepLabV3+, and SegFormer using images from the Sentinel-2 satellite, evaluating their performance through metrics such as Intersection over Union (IoU) and Dice coefficient. The results indicate that SegFormer achieved the highest accuracy in cloud detection but with a longer inference time, while U-Net demonstrated a balance between accuracy and efficiency, and DeepLabV3+ stood out for its shorter processing time but lower performance. The study highlights the relevance of neural networks in remote sensing and indicates the need for model optimization.*

1. Introduction

With the continuous advancement of satellite remote sensing, the amount of earth observation data has reached astonishing proportions [Janga et al. 2023], providing valuable information on land use and ecosystem changes [Gui et al. 2024]. Due to its ability to identify relevant characteristics, this technology is widely used in various applications, such as monitoring environmental indices [Santos et al. 2024], modeling deforestation [Ventura et al. 2023], as well as analyzing changes in land cover [Farnaz et al. 2025], with the aim of ensuring sustainability and efficient management of natural resources.

Among the satellite missions driving this progress is Sentinel-2, part of the European Space Agency's (ESA) *Copernicus* program[1]. Widely used in environmental monitoring [Santos et al. 2024, Farnaz et al. 2025], Sentinel-2 provides high-resolution, wide-coverage multispectral optical images, enabling detailed analysis of objects of interest [Misra et al. 2020]. However, despite the benefits offered by optical data from satellites such as Sentinel-2, the quality of this data is often affected by atmospheric conditions, such as the presence of clouds and cloud shadows, which reduce its usability [Domnich et al. 2021].

High cloud cover significantly compromises the accuracy of observations, altering the real values and introducing noise into the images due to the obstruction of pixels [Uchegbulam et al. 2021]. This problem affects various fields in remote sensing, such as the identification of areas affected by seismic disasters [Robinson et al. 2019], the monitoring of agricultural areas [Ozdogan et al. 2024] and the analysis of hydrographic quality [Langhorst et al. 2024], among others. Therefore, the accurate detection of clouds in images needs to be considered in order to minimize misinterpretations and, consequently, avoid incorrect conclusions in various applications [Luotamo et al. 2020].

In this context, the use of computer vision models for cloud segmentation in multispectral images has become a growing trend in recent years [Xu et al. 2024, Xu et al. 2025], especially due to their ability to extract spatial characteristics during image processing [Li et al. 2021]. Thus, deep learning models such as DeepLabV3Plus [Chen et al. 2018], U-Net [Ronneberger et al. 2015] and SegFormer [Xie et al. 2021] have been increasingly applied for this purpose [Wang et al. 2024], bringing significant contributions to cloud segmentation.

In this context, this study carries out a comparative analysis of three different deep learning architectures in the literature, with the aim of evaluating their performance in segmenting clouds in Sentinel-2 multispectral images. The models evaluated include DeepLabV3Plus, U-Net and SegFormer. The main contributions of this study include: (i) providing a detailed comparative analysis of the performance metrics and inference time of the specified models, and (ii) advancing cloud segmentation research in multispectral images, serving as a basis for future investigations.

The work is divided into the following parts: Section 2 details the selection and preparation of the data, and then the execution and evaluation of the model; Section 3 presents comparisons of the performance and inference metrics of the models, and their relationship with the differences in architecture; and finally, Section 4 summarizes the work carried out and its results, pointing out limitations of the work and future approaches to research.

2. Methodology

The methodology has been organized into three sections. In the sub-section 2.1, the **Data Selection and Preparation** is presented. In the sub-section 2.2, the compared computer vision architectures are detailed, with emphasis on their characteristics and applications in cloud detection in satellite images. Finally, the sub-section 2.3 discusses **Training and Evaluation**, covering the configurations adopted and the evaluation metrics. Fig-

[1]https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2

ure 1 details the order of the methodological procedures and the representations of the architectures used.

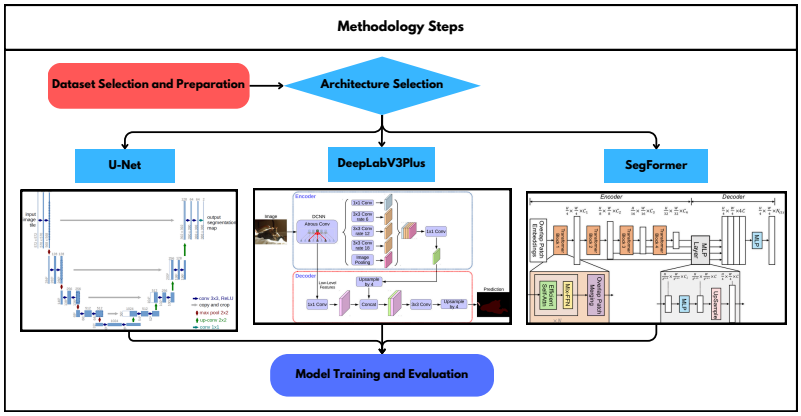

Figure 1. Flowchart of this article's methodology.

2.1. Data Selection and Preparation

The dataset *Sentinel-2 Cloud Mask Catalogue* [Francis et al. 2020] contains 513 Sentinel-2 satellite images, randomly sampled globally throughout 2018. The images have a spatial resolution of 20 meters and a size of 1022 × 1022 pixels, processed at the L1C level, i.e. geometrically corrected and expressed in reflectance at the top of the atmosphere, ideal for cloud analysis. The data was annotated semi-automatically by the IRIS tool[2], with dynamic Random Forest training by LightGBM (Gradient Boosting Machine), which both speeds up the annotation process by iteratively improving its predictions and allows the expert to make adjustments to the evaluation when necessary.

The authors provide three classes of segmentation mask: clean, cloud and cloud shadow, which correspond to 45.95%, 52.57% and 1.46% of the masks present in the dataset, respectively. However, it should be noted that not all the images have annotations for cloud shadows, with 86 images lacking this information. It was therefore decided to remove the cloud shadow class, since reducing the dataset due to the absence of this class or including images without the respective masks could compromise the models' ability to understand this category. The new proportion obtained corresponds to 46.64% clean pixels and 53.36% classified as cloud.

We used the RGB bands, commonly used for cloud segmentation [Domnich et al. 2021], band 10, specific for detecting cirrus clouds, and bands 11 and 12, used to differentiate cloud and snow pixels, which are relevant due to the presence of these elements in the images, which could hinder training due to their similarity in the visible spectrum[3]. To divide the data into training, test and validation, we considered a stratified 70%/15%/15% split based on the percentage of cloud present in each image, seeking better representativeness between the sets.

2.2. Computer Vision Architectures

Deep learning architectures, especially those based on Convolutional Neural Networks (CNNs), have played a key role in semantic image segmentation tasks. Classic mod-

[2]https://github.com/ESA-PhiLab/iris

[3]More information on: https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/bands/

els, such as U-Net and DeepLabV3Plus, which use architectures based on fully convolutional networks (FCN), are recognized for their high performance in medical applications [Ronneberger et al. 2015, Saifullah et al. 2025], which require extreme precision, and environmental applications [Wang et al. 2024].

Complementarily, with the recent rise of Transformer-based architectures, which are capable of capturing the global context of images, models such as SegFormer have shown significant advances in segmentation [Xie et al. 2021]. In medical tasks, SegFormer has already been compared with U-Net, obtaining performance equal to or better than the same [Sourget et al. 2023].

This subsection presents the three neural network architectures used for comparison: U-Net, DeepLabV3Plus and SegFormer. Each of these architectures will be detailed in the subsections 2.2.1, 2.2.2 and 2.2.3, discussing how they work and why they were chosen for the task of semantic cloud segmentation.

2.2.1. U-Net

U-Net is a semantic segmentation architecture based on a *encoder-decoder* approach. In the *encoder* phase, the image's spatial information is compressed, following the traditional structure of a FCN, consisting of a repeated sequence of two convolutions followed by the application of the *rectified linear unit* (ReLU) activation function and a *Max Pooling* operation. In the *decoder* phase, the original image size is restored using transposed convolutions instead of traditional convolutions. The main feature of U-Net is the concatenation of the compressed information from the *encoder* phase with the decompressed information in the *decoder* phase, forming so-called *skip connections*, which allow important spatial details to be preserved [Ronneberger et al. 2015].

The extensive systematic review carried out by [Wang et al. 2024] compares various architectures in the context of cloud detection in different satellite datasets, highlighting that U-Net-based models outperformed traditional models, such as those based on *Random Forest* and Fmask. In addition, [Domnich et al. 2021] developed KappaMask, a cloud detection model based on U-Net, which outperformed popular models used in Sentinel-2, demonstrating the architecture's great potential for this task.

2.2.2. DeepLabV3Plus

The DeepLabV3Plus architecture is a direct evolution of DeepLabV3, adding a dilated convolution mechanism to the decoding operation of the [Chen et al. 2018] architecture. Like U-net, the architecture behaves similarly to FCNs in the encoder phase, but incorporates an additional phase called *Spatial Pyramid Pooling* (SPP). SPP applies filters of varying sizes with dilation at different spatial divisions of the image, allowing the network to capture information at different scales of spatial resolution.

In the context of cloud segmentation, previous studies, such as the one by [Liu et al. 2019], carried out a comparative analysis between DeepLabV3Plus, Sen2Cor and FCNs, highlighting the superiority of DeepLabV3Plus in various performance metrics. In addition, a recent study [He et al. 2024] compared popular architectures such

as U-Net, ResNet and DeepLabV3 with DeepLabV3Plus on a Sentinel-2 dataset with semi-automatically annotated clouds, as a result of which DeepLabV3Plus excelled in practically all metrics.

2.2.3. SegFormer

SegFormer is an architecture based on *encoder-decoder* that doesn't use convolutions, but rather layers of hierarchical *transformers* and layers of *multilayer perceptrons* (MLPs). Initially, the image is divided into smaller portions and then passed through a *encoder* called *Mix Vision Transformers (MiT)*. The main feature of SegFormer is its efficiency, which is achieved by reducing the complexity of the *decoder*, using only MLP layers, which results in lower computational cost without sacrificing too much performance [Xie et al. 2021].

Previous studies in cloud segmentation, such as [Choi et al. 2024], highlight the architecture's high performance, which obtained higher *F1-Score* values than U-Net variants, such as HRNet and Unet3+, and customized models based on *encoder-decoder* architectures, such as *Cloud Detection-Fusing Multi-Scale Spectral and Spatial Features* (CD-FM3SF). In addition, [Wang et al. 2024] compared the architecture with others based on attention mechanisms, obtaining competitive performance in metrics such as *intersection-over-union* (IoU).

2.3. Training and Evaluation

The machine configuration used consisted of an Nvidia A100, with 40 GB of VRAM memory and 100 GB of RAM memory, which provided sufficient resources to use a batch size of 8. The library *segmentation models pytorch*[4] was used to implement the models. For DeepLabV3Plus, we opted for the *encoder* MobileNetV2, with two million parameters, as suggested in the original proposal. For U-Net, as there was no specific suggestion in the initial proposal, the same *encoder* was used to facilitate comparison between the architectures. In the case of SegFormer, the *encoder* MIT-B0 was chosen, with three million parameters, as specified in the original proposal, ensuring a fair comparison of the architecture in its standard configuration [Xie et al. 2021].

Both *encoders* were pre-trained on the ImageNet[5] dataset to improve the initialization of the weights, which enables faster convergence and more effective weights for the [Koonce 2021] task. In addition, *Dice Loss* was chosen as the loss function, as this function directly optimizes the overlap of the segmented masks and is widely used for [Azad et al. 2023] segmentation tasks.

To monitor training, the *PyTorch Lightning*[6] library was used, with implementations of techniques such as *EarlyStopping* and Reducing the Learning Rate at the Plateau (*ReduceLrOnPlateau*). The *EarlyStopping* technique was configured to monitor the loss in the validation set, with a stop criterion after 10 epochs without improvement. On the other hand, *ReduceLrOnPlateau* was configured to monitor the validation loss and reduce the learning rate, initially at 10^{-4}, every 5 epochs, with a reduction factor of 50%,

[4]https://smp.readthedocs.io/en/latest/

[5]https://www.image-net.org/

[6]https://lightning.ai/docs/pytorch/stable/

in order to prevent the gradient from losing optimal values in the validation loss. Both configurations were defined empirically, allowing training to be extended and possible improvements in model performance to be observed. In addition, the best model obtained at each training session, based on the minimum value of the validation loss, was saved.

Finally, the metrics used were ***Intersection over Union*** (IoU), which assesses the ratio between the segmentation inferred by the model and the actual segmentation, with higher values being desirable as it provides a clear measure of the quality of the global cloud segmentation; **precision** and **recall**, two classic and complementary metrics in which higher values indicate, respectively, better ability of the model to avoid false positives and greater efficiency in identifying all cloud instances; the coefficient ***Dice***, which measures the similarity between the inferred segmentation and the real segmentation, relevant in scenarios where the data set is unbalanced [Azad et al. 2023], since high values reflect greater agreement between the segmentations; and finally, the average inference time, calculated in seconds, based on 10 random samples from the test set.

3. Results and Discussion

Table 1 shows the metrics obtained for each model, with the best metrics highlighted in bold. A similar performance is observed between SegFormer and U-Net, with distinct advantages for each model. SegFormer had the highest IoU (0.85), which shows a more consistent intersection in cloud segmentation with only 10 epochs to reach its best performance. It also had the highest recall (0.90), which indicates a greater ability to correctly identify cloud regions. Despite requiring the highest number of epochs (26), U-Net, obtained the highest *Dice* coefficient (0.91), indicating greater similarity between the predicted and real masks, as well as greater accuracy (0.94), which demonstrates its superior ability to avoid false positives. DeepLabV3Plus, on the other hand, lagged behind in all metrics despite requiring 19 epochs to reach its best model and presenting values close to those of the other architectures, suggesting less accurate segmentation using this approach.

Architecture	*Dice*	IoU	Recall	Precision	Epoch
U-Net	**0.9079**	0.8317	0.8761	**0.9427**	26
DeepLabV3+	0.8872	0.8106	0.8702	0.9228	19
SegFormer	0.9043	**0.8534**	**0.9080**	0.9307	**10**

Table 1. Comparing the Performance of Architectures

Table 2 shows the inference times together with the relative distance between each of the architectures. It can be seen that, despite the high performance shown by SegFormer, it had the longest inference time (1.12s) and the slowest relative distance between both architectures, with an increase of +72.3% compared to U-Net and +63.4% compared to DeepLabV3+, a limitation in terms of efficiency.

On the other hand, U-Net showed greater balance between the metrics, delivering a reasonable inference time (0.65s) and moderate percentage distances, with +36.9% compared to DeepLabV3+ and -41.96% compared to SegFormer. DeepLabV3+, on the other hand, had the shortest inference time (0.41s) and the fastest speed, with a reduction

of -58.5% and -173.2%, respectively, compared to U-Net and SegFormer, which can be an advantage in applications that require greater speed. The choice of the ideal model depends on the balance between these factors.

	Inference Time (s)	U-Net	DeepLabV3+	SegFormer
U-Net	0.65	0%	-58.5%	+72.3%
DeepLabV3+	0.41	+36.9%	0%	+63.4%
SegFormer	1.12	-41.96%	-173.2%	0%

Table 2. Percentage distance between the model's inference times, including the original inference time. Positive values indicate that the model in the column is slower than the model in the row.

Figure 2 shows a comparison between the masks generated by each model on two randomly selected images, based on the percentage of cloud present in the images. To make it easier to see the masks, a threshold was set between 10% and 50%. It can be seen that SegFormer produced accurate masks in both cases, demonstrating a greater ability to identify different cloud trails.

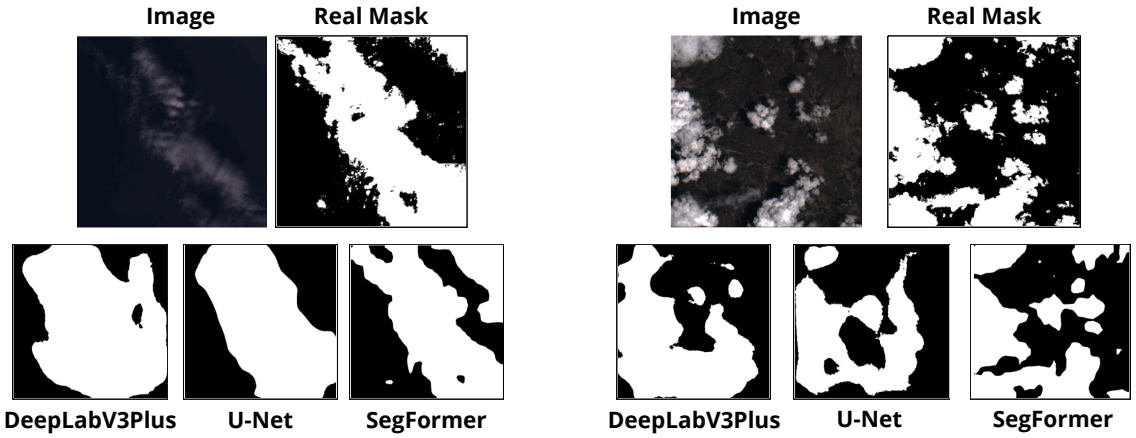

Figure 2. Visual Comparison of Masks Obtained with Different Models

The performance of U-Net, which demonstrated good results in terms of precision and DICE coefficient in this study, is widely corroborated by the literature. [Wang et al. 2024] highlight that U-Net-based models often outperform traditional approaches and remain a robust and effective choice for cloud segmentation tasks. [Domnich et al. 2021], for instance, developed KappaMask, a U-Net-based cloud detection model for Sentinel-2, which was shown to surpass other popular models, thereby reinforcing the potential of this architecture.

Regarding DeepLabV3+, the obtained results indicate a consistent performance, albeit slightly inferior to that of U-Net and SegFormer, considering the adopted dataset and experimental configuration. Interestingly, other studies, such as [Liu et al. 2019] and [He et al. 2024], both focused on cloud segmentation in Sentinel-2 imagery, highlight the superiority of DeepLabV3+ across various performance metrics. This apparent divergence suggests that factors such as dataset specificity, pre-processing strategies, encoder choices, and training hyperparameters can exert a significant influence on the relative performance of the architectures.

These results are in line with trends observed in the literature, with models based on *Transformers* being a direct evolution of the old fully convolutional architectures, surpassing them in various segmentation tasks [Liu et al. 2023, Wang et al. 2024]. This is because these blocks make it possible to capture global contexts in images, while preserving local details, which is relevant for large, complex images, such as the multispectral ones present in the dataset used.

4. Conclusion

This comparative study evaluated different Deep Learning approaches for cloud detection in multispectral images from the Sentinel-2 satellite. The results showed that the SegFormer-based model performed best in the IoU and recall metrics, proving superior at identifying all cloud areas, although it had a longer inference time. Additionally, U-Net showed better ability to avoid false positives and a high similarity to real cloud regions, but may not be as efficient as SegFormer in identifying all cloud regions, despite showing an advantage with a shorter inference time and balanced metrics.

However, some limitations must be considered. Firstly, the data set used refers to a specific period, which may limit the generalization of the results acquired, especially in terms of time scale. In addition, the low density of training data may have compromised the performance of the SegFormer architecture, which, due to the use of Transformers, requires a large amount of data to reach its maximum potential. Also, as the evaluation of the models depends on the utilized similarity metrics, it is challenging to determine the best architecture, as it depends on the priorities and resources available to the researcher.

Future investigations could explore enhancements to the evaluated models, including the comparison of different encoders and their impact on performance when combined with the aforementioned architectures. It is also relevant to explore pre-processing techniques, such as image histogram normalization to mitigate the effect of outliers, and the inclusion of more spectral bands, assessing how these strategies influence segmentation quality. Furthermore, it is important to consider expanding the set of metrics used in model evaluation, incorporating measures such as Overall Accuracy (OA), F1-Score, mean Intersection over Union (mIoU), and mean Accuracy (mA) [Wang et al. 2024], enabling a more comprehensive analysis of both pixel-wise segmentation and overall model performance.

Regarding the dataset utilized, plans include its expansion and diversification, either through the incorporation of additional data with more complete annotations or the application of data augmentation techniques. Consequently, it is planned to include the analysis of cloud shadows, as well as to overcome the limitation of low data density, thereby encompassing an evaluation with greater temporal, spatial, and spectral variability. Finally, to serve as a basis for further research, a code repository was created on the GitHub[7] platform. This initiative aims to facilitate replication of the experiments and contribute to the academic community.

References

Azad, R., Heidary, M., Yilmaz, K., Hüttemann, M., Karimijafarbigloo, S., Wu, Y., Schmeink, A., and Merhof, D. (2023). Loss functions in the era of semantic seg-

[7]https://github.com/LPO-LIIS/CloudDetectionSentinel-2

mentation: A survey and outlook. *arXiv preprint arXiv:2312.05391*.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.

Choi, J., Seo, D., Jung, J., Han, Y., Oh, J., and Lee, C. (2024). Cloud detection using a unet3+ model with a hybrid swin transformer and efficientnet (unet3+ ste) for very-high-resolution satellite imagery. *Remote Sensing*, 16(20):3880.

Domnich, M., Sünter, I., Trofimov, H., Wold, O., Harun, F., Kostiukhin, A., Järveoja, M., Veske, M., Tamm, T., Voormansik, K., et al. (2021). Kappamask: Ai-based cloudmask processor for sentinel-2. *Remote Sensing*, 13(20):4100.

Farnaz, Nuthammachot, N., Shabbir, R., and Iqbal, B. (2025). Pixel and region-oriented classification of sentinel-2 imagery to assess lulc dynamics and their climate impact in nowshera, pakistan. *Open Geosciences*, 17(1):20220745.

Francis, A., Mrziglod, J., Sidiropoulos, P., and Muller, J.-P. (2020). Sentinel-2 cloud mask catalogue.

Gui, S., Song, S., Qin, R., and Tang, Y. (2024). Remote sensing object detection in the deep learning era—a review. *Remote Sensing*, 16(2).

He, M., Zhang, J., He, Y., Zuo, X., and Gao, Z. (2024). Annotated dataset for training cloud segmentation neural networks using high-resolution satellite remote sensing imagery. *Remote Sensing*, 16(19):3682.

Janga, B., Asamani, G. P., Sun, Z., and Cristea, N. (2023). A review of practical ai for remote sensing in earth sciences. *Remote Sensing*, 15(16).

Koonce, B. (2021). *Convolutional Neural Networks with Swift for Tensorflow*. Springer.

Langhorst, T., Andreadis, K. M., and Allen, G. H. (2024). Global cloud biases in optical satellite remote sensing of rivers. *Geophysical Research Letters*, 51(16):e2024GL110085.

Li, J., Wu, Z., Hu, Z., Jian, C., Luo, S., Mou, L., Zhu, X. X., and Molinier, M. (2021). A lightweight deep learning-based cloud detection method for sentinel-2a imagery fusing multiscale spectral and spatial features. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–19.

Liu, C.-C., Zhang, Y.-C., Chen, P.-Y., Lai, C.-C., Chen, Y.-H., Cheng, J.-H., and Ko, M.-H. (2019). Clouds classification from sentinel-2 imagery with deep residual learning and semantic image segmentation. *Remote Sensing*, 11(2):119.

Liu, Y., Zhang, Y., Wang, Y., Hou, F., Yuan, J., Tian, J., Zhang, Y., Shi, Z., Fan, J., and He, Z. (2023). A survey of visual transformers. *IEEE Transactions on Neural Networks and Learning Systems*.

Luotamo, M., Metsämäki, S., and Klami, A. (2020). Multiscale cloud detection in remote sensing images using a dual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):4972–4983.

Misra, G., Cawkwell, F., and Wingler, A. (2020). Status of phenological research using sentinel-2 data: A review. *Remote Sensing*, 12(17).

Ozdogan, M., Wang, S., Ghose, D., Fraga, E. P., Fernandes, A. M., and Varela, G. (2024). Field-scale rice area and yield mapping in sri lanka with optical remote sensing and limited training data. *Available at SSRN 4940849*.

Robinson, T. R., Rosser, N., and Walters, R. J. (2019). The spatial and temporal influence of cloud cover on satellite-based emergency mapping of earthquake disasters. *Scientific reports*, 9(1):12455.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.

Saifullah, S., Dreżewski, R., and Yudhana, A. (2025). Advanced brain tumor segmentation using deeplabv3plus with xception encoder on a multi-class mr image dataset. *Multimedia Tools and Applications*, pages 1–22.

Santos, M., Paula, M., and Souza, V. (2024). Extração de séries espaço-temporais a partir de um cubo de dados geoespacial em benefício da cafeicultura mineira. In *Anais do XV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais*, pages 211–214, Porto Alegre, RS, Brasil. SBC.

Sourget, T., Hasany, S. N., Mériaudeau, F., and Petitjean, C. (2023). Can segformer be a true competitor to u-net for medical image segmentation? In *Annual Conference on Medical Image Understanding and Analysis*, pages 111–118. Springer.

Uchegbulam, O., Ameloko, A., and Omo-Irabor, O. (2021). Effect of cloud cover on land use land cover dynamics using remotely sensed data of western niger delta, nigeria. *Journal of Applied Sciences and Environmental Management*, 25(5):799–804.

Ventura, R., Musis, C., Ventura, T., and II, A. C. (2023). Rsdd: um conjunto de dados para modelagem de desmatamentos. In *Anais do XIV Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais*, pages 159–162, Porto Alegre, RS, Brasil. SBC.

Wang, Z., Zhao, L., Meng, J., Han, Y., Li, X., Jiang, R., Chen, J., and Li, H. (2024). Deep learning-based cloud detection for optical remote sensing images: A survey. *Remote Sensing*, 16(23):4583.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090.

Xu, K., Wang, W., Deng, X., Wang, A., Wu, B., and Jia, Z. (2024). Transga-net: Integration transformer with gradient-aware feature aggregation for accurate cloud detection in remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5.

Xu, X., He, W., Xia, Y., Zhang, H., Wu, Y., Jiang, Z., and Hu, T. (2025). Tanet: Thin cloud-aware network for cloud detection in optical remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–16.