Mapping stains on flat roofs using semantic segmentation based on deep learning

Lara Monalisa Alves dos Santos, Leonardo Rabero Lescano, Gabriel Toshio Hirokawa Higa, Vanda Alice Garcia Zanoni, Lenildo Santos da Silva, César Iván Álvarez, Hemerson Pistori

Angaben zur Veröffentlichung / Publication details:

Alves dos Santos, Lara Monalisa, Leonardo Rabero Lescano, Gabriel Toshio Hirokawa Higa, Vanda Alice Garcia Zanoni, Lenildo Santos da Silva, César Iván Álvarez, and Hemerson Pistori. 2025. "Mapping stains on flat roofs using semantic segmentation based on deep learning." *Case Studies in Construction Materials* 22: e04106. https://doi.org/10.1016/j.cscm.2024.e04106.

Case Studies in Construction Materials 22 (2025) e04106

Contents lists available at ScienceDirect

Case Studies in Construction Materials

journal homepage: www.elsevier.com/locate/cscm

Case study

Mapping stains on flat roofs using semantic segmentation based on deep learning

Lara Monalisa Alves dos Santos [a,*], Leonardo Rabero Lescano [b], Gabriel Toshio Hirokawa Higa [b], Vanda Alice Garcia Zanoni [a], Lenildo Santos da Silva [a], Cesar Ivan Alvarez [c,d], Hemerson Pistori [b,e]

[a] University of Brasilia, Distrito Federal, Brazil
[b] Dom Bosco Catholic University, Mato Grosso do Sul, Brazil
[c] Centre for Climate Resilience, University of Augsburg, Universitätsstrasse 12a, Augsburg 86159, Germany
[d] Environmental Research Group for Sustainable Development (GIADES), Salesian Polytechnic University, Rumichaca y Moran Valverde, Quito, Ecuador
[e] Federal University of Mato Grosso do Sul, Mato Grosso do Sul, Brazil

ARTICLE INFO

Keywords:
Maintenance
Building inspection
Computer Vision
CNN – FCN and DeepLabV3
Transformer – SegFormer

ABSTRACT

Moisture stains indicate ongoing degradation processes and may reveal areas of the roof slab where water infiltration occurs, compromising the performance and durability of the building system. During inspections of roofing systems, an inspector's field of vision differs from that of drones during overflights. As a result, traditional inspections might not always detect the presence and severity of stains, making maintenance on flat roofs a complex task. In this context, this experimental study aims to analyze deep learning-based semantic segmentation with images obtained from drones to map and monitor damp patches during automated building inspections of flat roof systems. The research tested two convolutional neural networks for semantic segmentation: the Fully Convolutional Network (FCN) with a ResNet50 backbone and DeepLabV3 with a ResNet101 backbone, as well as a transformer-based deep artificial neural network called SegFormer with a MiT-B1 backbone. We evaluated three optimizers for each model—Adam, Adagrad, and SGD—along with learning rates of 1e-2, 1e-3, and 1e-4. The models were compared using four performance metrics. The FCN, optimized with Adagrad at a learning rate of 1e-2, showed the best results. The average metrics obtained in this case were as follows: precision: 79.69 %, recall: 67.81 %, F-score: 73.09 %, and Intersection over Union (IoU): 57.70 %.

1. Introduction

Waterproofing systems are important in ensuring the watertightness of flat roofs, helping them maintain their performance under various exposure conditions. After rainfall, it's common for puddles and damp stains to appear on the waterproofing layer and its mechanical protection. The main causes of damage include surface irregularities and inadequate slopes that hinder proper rainwater runoff. When applying the waterproofing system, the orientation of the blanket strips—whether longitudinally or transversely positioned relative to the drainage pipes—can lead to issues with water pooling due to the overlaps and seams. Additionally, any bumps in

* Corresponding author.
E-mail address: 200054104@aluno.unb.br (L.M.A. Santos).

https://doi.org/10.1016/j.cscm.2024.e04106
Received 3 August 2024; Received in revised form 28 October 2024; Accepted 9 December 2024
Available online 18 December 2024

the blankets and maintenance repairs can result in the accumulation of dirt, biological growth, and the spread of invasive vegetation, contributing to the development of moisture stains.

Most stains are harmful because they affect the aesthetic aspect in the initial phase, but if they remain, they worsen other pathological manifestations [1]. Stains may indicate regions where water infiltration occurs, accelerating degradation mechanisms and resulting in the loss of habitability conditions. Sometimes, to find the origin of the infiltrations, it is necessary to remove layers of coating, mortar, and protection [2–4].

In this study, moisture stains are treated as indicators of points or regions on the roof slab where infiltrations occur, compromising the performance of the building system. Moisture stains on the flat roof can vary in size, color, tone, and severity; they may dry out depending on exposure to sunlight and wind. However, the stains indicate progressive degradation processes that affect the durability of the roofing components.

Maintenance actions in buildings mitigate the worsening of damage. However, traditional inspection methods require the involvement of experienced professionals to evaluate the systems' performance [5]. In most cases, inspections become complex due to the roof's size and inaccessibility for visual analysis [6,7].

The detection of pathological manifestations carried out by specialists in manual processes is time-consuming and depends on weather conditions [8–10]. In building inspections of roofing systems, the inspector's field of vision responsible for the inspection differs from the field of vision of the drone's cameras during the overflight. For this reason, traditional inspections do not always identify the presence and severity of stains across the entire roof surface.

In recent years, methods for automated building inspections have advanced in detecting pathological manifestations. Despite the rapid progress of digital technologies and data processing types based on deep learning (DL), there is still a vast area for research and development [11]. In this context, Cao [12] considers that damage to the surfaces of architectural components receives little attention since studies on structural health monitoring (SHM) using DL, focused on detecting cracks and structural deformations, predominate.

Perez, Tah, and Mosavi [7] used convolutional neural networks (CNNs) to classify mold, stains, and surface deterioration images. The VGG16 model (Convolutional Network for Classification and Detection) was used and evaluated by comparison with the ResNet50 and Inception models. Class activation mapping was used to locate the defects. The authors considered the nature of the damage and the surrounding environment as challenges since mold in houses can be black, brown, green, olive green, gray, blue, white, yellow, or even pink. In addition, stains and paint deterioration do not have a defined shape, size, or color, and their physical characteristics are strongly influenced by the location (walls, ceilings, corners), the background (paint color, wall texture and material, etc.), and light intensity.

Using devices for capturing high-quality images, such as smartphones and drones equipped with cameras, results in better object detection accuracy [13]. Additionally, convolutional neural networks allow for advanced visual analysis and understanding [14]. Several studies have focused on the use of semantic image segmentation [15] for the detection of pathological manifestations in buildings [16,17]. Semantic segmentation assigns a category label to each pixel in an image, providing category information at the pixel level, and this has been applied to damage detection [18].

The use of semantic segmentation associated with drone-captured images, with annotations of roof images, has shown significant results. Yundin, Naumov, and Patrakova [9] developed software for recognizing irregularities in waterproof blankets on flat roofs. Using a semantic segmentation algorithm makes it possible to detect irregularities that cause water stagnation on flat roofs after rain. The software is fed with drone-obtained images, automates the recognition of water pooling, and prepares binary masks. The approach based on Fully Convolutional Networks (FCN) demonstrated results with a dice score of up to 0.78 in the proposed task. The authors warn about the issues related to sensitivity to shadows, changes in object lighting, and different types of roofs, which require significant complexity in image recognition algorithms.

Few studies have been found in the literature dealing with semantic segmentation in roofs. Specifically, when it comes to stains on flat roofs, the research by Yudin, Naumov, and Patrakova [9] stands out. They use annotations of the region of interest in binary masks and apply FCN as the network architecture. This motivates us to compare other deep learning-based semantic segmentation models to deepen and master their application.

This experimental study aims to analyze the application of deep learning-based semantic segmentation using drone-acquired images to map and monitor moisture stains during automated inspections of flat roof systems. Two types of convolutional neural networks for semantic segmentation were tested: Fully Convolutional Networks (FCN) and DeepLabV3, along with a transformer-based neural network called SegFormer. These models were compared using four performance metrics.

Automatic recognition of pathological manifestations improves data collection, processing, and analysis efficiency and speed. This, in turn, ensures durability and extends the service life of buildings. Therefore, the experiments conducted in this study contribute to the advancement of automated techniques for detecting stains on flat roofs, integrating digital technologies into building inspection procedures.

2. Overview

Maintenance management, especially building inspection processes, incorporates approaches or strategies based on machine learning (ML) to replace traditional methods. New digital technologies enhance integrated methodologies that significantly transform construction tasks, making them safer, more efficient, and adaptable to contemporary needs [19].

In performance evaluations aimed at decision-making in preventive maintenance, artificial intelligence (AI) approaches are increasingly favored over those based on physical or statistical models [20]. Data-driven pattern recognition enables predictive models that enhance decision-making processes [21,22]. However, the integration of digital technologies with robust and efficient machine

learning (ML) predictive models remains an area for further research and experimentation [11,19].

Machine learning strategies can improve the accuracy and reliability of engineering prediction models, especially when combined with hybrid approaches, in contexts where traditional methods may be insufficient. [19] used Gradient-Boosting algorithms to forecast the shear strength for SFRC-Slender Beams without Stirrups (SFRC-SBWS) associated with traditional engineering calculation models. The authors emphasize the relevance of machine learning techniques in making structural designs more reliable, safe, and resilient.

Traditional ML algorithms, such as Gradient Boosting and its variations (e.g., Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and CatBoost), can be very effective in classification and regression tasks with tabular data [23,24]. However, they are not specifically designed for image data. These algorithms can generally be used in computer vision, which requires feature extraction and selection through image preprocessing techniques. According to [25], traditional algorithms have certain advantages when applied to the construction sector, as they require fewer labeled data than deep learning algorithms and are more interpretable than deep learning models.

On the other hand, in recent years, CNNs have distinguished themselves by their superior performance compared to traditional algorithms in several computer vision tasks [25]. Models such as ResNet, DenseNet, and EfficientNet have demonstrated high performance and generalization capabilities in computer vision benchmarks. CNNs are generally more effective due to their ability to extract complex spatial features automatically, boosting their performance in tasks such as image recognition and classification, object detection, and semantic segmentation [26].

Fully Convolutional Networks (FCNs) are a class of neural networks designed for dense prediction tasks, such as semantic segmentation. They use convolutional layers to process input images of any size and return corresponding outputs. In addition, they combine semantic information from deep and wide layers with visual details from shallow and fine layers, which enhances the accuracy and detail of the segmentations [27].

The DeepLabV3 model differs from previous DeepLab versions by incorporating atrous (or dilated) convolutions in a cascade, allowing it to capture information at multiple scales without increasing the number of parameters. Additionally, it uses atrous spatial pyramid pooling (ASPP), which applies atrous convolutions with different rates in parallel to capture information at various scales, improving segmentation accuracy [28].

An emerging type of architecture that has gained prominence is the Transformer network, a model widely used in tasks such as natural language processing. This is due to its ability to process data in parallel and capture long-range dependencies, which differs from the approach of traditional CNNs, which use local convolutions to capture information in neighboring regions and process the image sequentially [29]. Furthermore, Transformers generalize better across different datasets and segmentation scenarios, making them a promising choice for image segmentation tasks due to these characteristics, which overcome some limitations of traditional CNN models.

SegFormer is an architecture that combines Transformers with multilayer perceptron (MLP) decoders to achieve high efficiency and accuracy in semantic segmentation tasks. Its Transformer encoder generates multiscale representations, while the simple MLP decoder integrates local and global attention. In addition, it demonstrates robustness in zero-shot scenarios, making it an excellent choice for tasks involving critical security applications [30].

For optimizing neural networks, the algorithms SGD (Stochastic Gradient Descent), Adam (Adaptive Moment Estimation), and Adagrad (Adaptive Gradient) are widely used due to their effectiveness in adjusting the weights of neural networks during training. Among the strategies used to improve the optimization of neural networks, their performance approaches range from hyperparameter tunings, such as learning rate and batch size, to more advanced regularization techniques, such as dropout and L2 regularization,

Fig. 1. Location of the buildings at the Headquarters of the Brazilian Army (QGEx) in Brasília, capital of Brazil.

which mitigate overfitting. Transfer learning is also commonly applied, enabling pre-trained models to adapt quickly to new tasks with less data by leveraging prior knowledge [31].

The literature review supported the selection of deep learning algorithms and defined the approach of this research. Considering their application in image segmentation in civil construction [25], it is possible to assume that the techniques used represent the state of the art in computer vision experiments. The motivation for the case study developed is justified by the need to advance the application of digital technologies and machine learning in information management for preventive maintenance tasks.

Innovative neural networks based on Transformers present specific challenges that require benchmarking. Since their conception in Natural Language Processing (NLP), Transformer neural networks have often outperformed CNNs. With Transformers now being applied to other fields, such as computer vision, additional experiments are essential to evaluate their effectiveness and adaptability to these new domains, especially complex environments such as flat roofs.

3. Research methodology

3.1. Image dataset

For the proposed experiment, data were collected from 9 buildings located at the Brazilian Army Headquarters (QGEx) in Brasília, the capital of Brazil, as shown in Fig. 1. The dimensions of the buildings' flat roofs vary between 200.40 m x 12.50 m and 250.80 m x 14.90 m.

The images were captured using the Mavic Pro drone manufactured by DJI - DÀ-JIĀNG Innovations Science and Technology [32], whose camera parameters are shown in Table 1.

To capture images of the presence of stains on the waterproofed slabs of the flat roofs, three flight plans were made using DJI PILOT [19] software, coordinated by the Geographic Positioning System (GPS). The three flights were of the cross-type, in both directions (Fig. 2), with a frontal and lateral overlap rate of 60 %. Following the protocol proposed by Santos et al. [3], the shortest distance between the flat roof and the drone, with heights of 22 and 35 m, was considered to guarantee the highest precision and detail in acquiring the images. In addition, the altitude of the flights was determined considering local conditions and existing barriers. Variations in positions and configurations were avoided to maintain uniformity in the images.

The flight plans took place on different days. The first flight was carried out on December 8, 2021, at 10:00 AM, on a sunny day with a temperature of 19.7 degrees and a wind speed of 0.5 m/s. The second flight was carried out on March 8, 2022, at 2:00 PM, on a sunny day with a temperature of 24.4 degrees and a wind speed of 3.6 m/s. The two flight plans provided 766 images acquired by the UAV Mavic Pro. Of this total, 126 images containing light and dark stains were selected for annotation.

After a period of rain, water pooling can occur, which may dry and leave light stains. This is a natural process due to the roof's exposure conditions and the fading of construction materials. Fig. 3 illustrates the light stains, which are not of interest in this study. Therefore, they were neither annotated nor considered segmentation targets.

The darker stains indicate the constant presence of moisture and stagnant water, associated with dirt and biological colonization, which accelerate the degradation process of the roof's waterproofing system or indicate possible infiltration. Fig. 4 exemplifies a stain characterized by a darker color, indicative of severity and a target of the semantic segmentation proposed for this research. A specialist identified the annotated dark stains in the images according to the standard procedure adopted in an on-site inspection after a rainy period.

After capturing the images and obtaining the reference sample, the 126 selected images were annotated for binary segmentation, considering stain and background pixels. The input color image was converted to the HSV color model, where H represents hue, S represents saturation, and V represents value. Smoothing was applied, and the resulting image was returned to the RGB color space. The ROBOFLOW platform [33] was used for annotation, which hosts public datasets and allows exporting in various formats, including the COCO (Microsoft Common Objects in Context) format, which was used for annotating the stains. The selected annotation technique was the polygon tool (Fig. 5), where a collection of coordinates is drawn around an image, avoiding capturing background information. Polygonal annotations (Fig. 6) allow more precise definitions of the labeled object, which can lead to improved performance and training results.

Table 2 presents the characteristics of the 126 images annotated on the ROBOFLOW platform [33]. Fig. 7 shows a graph of the resolutions of the annotated images. Fig. 8 depicts the distribution of the positions where the stains appear in the image dataset, ranging from 0 to 2501 annotations, considering the pixels contained in each polygon.

Table 1
DJI Mavic Pro UAV camera parameters.

Parameters	Value
Sensor	1/2.3" (CMOS), Effective Pixels: 12.35 M (Total Pixels: 12.71 M)
Lens	FOV 78.8° 26 mm (35 mm equivalent format) f/2.2 Distortion < 1.5 % focus from 0.5 m to ∞
ISO photo range	100–1600
Electronic shutter speed	8 s − 1/8000 s
Image dimensions	4000 × 3000 pixels

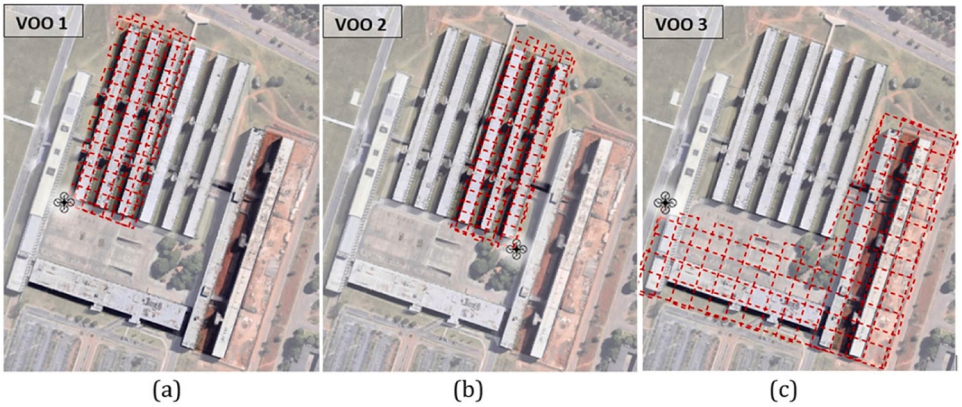

Fig. 2. Flight plans for image acquisition captured by the drone.

Fig. 3. Light moisture stains on the flat roof after rain.

Fig. 4. Reference area for annotating dark stains.

3.2. Experimental approach

In the experiment, three models were used: two convolutional neural networks, DeepLabV3 [28] with ResNet101 as the backbone and FCN [27] with ResNet50 as the backbone, and a transformer-based network, SegFormer [30] with MiT-B1 as the backbone. The Adam, Adagrad, and SGD algorithms were tested for neural network optimization for each architectural configuration. In each case, learning rates of 1e-2, 1e-3, and 1e-4 were evaluated, resulting in nine different optimization strategies for each neural network architecture.

The models were evaluated using a 10-fold cross-validation approach. 70 % of the 126 available images were allocated to the

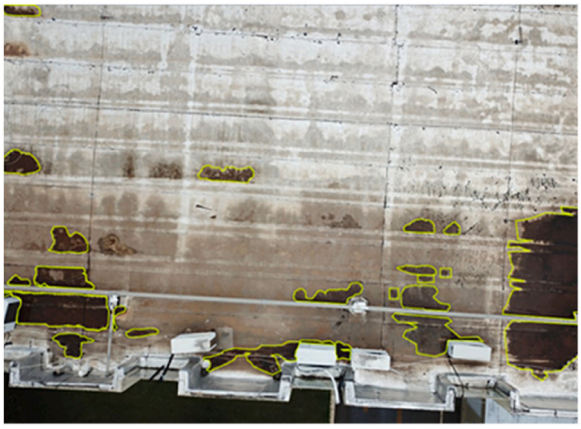

Fig. 5. Application of the Polygon Tool.

Fig. 6. Stain Annotation using ROBOFLOW.

Table 2
Description of the dataset on the ROBOFLOW platform: Class Balance - Stain.

Images	Annotations	Average Image Size	Median Image Ratio
126	2501	4.76 MP	4000 × 3000

training set, 20 % to the validation set, and 10 % to the test set for each fold. All images were resized to 512 by 512 pixels and subjected to the following data augmentation techniques during training: horizontal flip with a 50 % probability; random brightness contrast, where brightness and contrast were randomly adjusted by up to ± 20 %; and random sun flare. The images were then subjected to a series of additional modifications. These included simulating sunlight rays with a 20 % probability, introducing subtle fog with an intensity of 20 %, and applying random rotations between −25 and 25 degrees. The latter was implemented to allow the network to detect objects of interest in different orientations. The batch size was set to 6, and the loss function used was cross-entropy. The training was limited to a maximum of 1000 epochs, with early stopping implemented after 50 consecutive epochs without reducing the observed validation loss.

The performance of the models was then evaluated using four metrics (Table 3): Precision, Recall, F-score, and Intersection over Union (IoU). For a given image *i*, the number of true positives (TP), the number of false positives (FP), the number of true negatives (TN), and the number of false negatives (FN) were calculated. Mask is the manually annotated mask, and Prediction is the prediction given by the neural network. Finally, the final score in the test set for each fold was given as an average across all images. For a given image *i*, P*i*, R*i*, F1*i*, and IoU*i* were considered, as shown in Eq. 1, 3, 5, and 7. For a given class (in this case, stain), the metrics were defined as P, R, F, and IoU, where N is the number of test images, as shown in Eq. 2, 4, 6, and 8.

For visualization of the results, colors were used to represent the pixel-level segmentation for each case. Green was used for true positives (TP), red for false positives (FP), and blue for false negatives (FN). An example is shown in Fig. 9. A comparison is made between each pixel in the segmented image and the reference mask. If a pixel correctly identifies as belonging to a given spot, it is counted as a true positive (TP). If the pixel is incorrectly classified as part of a spot, it is counted as a false positive (FP). Pixels in the

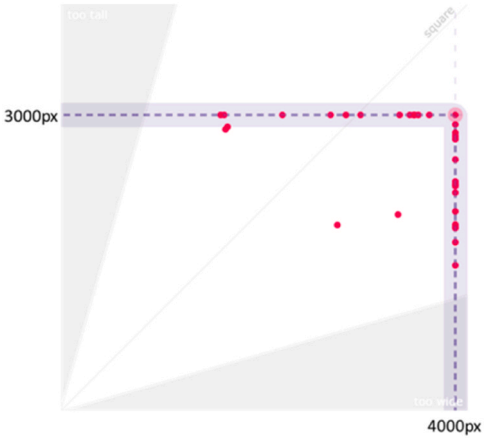

Fig. 7. Information on the dimensions of the annotated images.

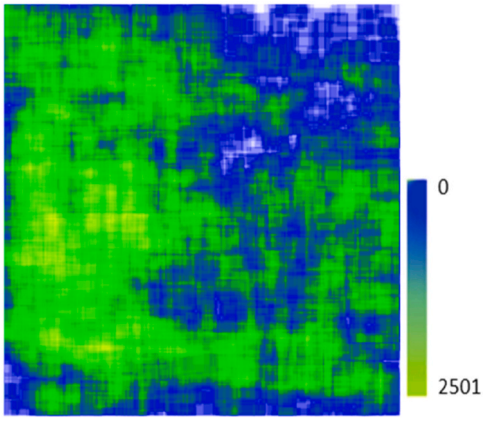

Fig. 8. Annotation heatmap.

Table 3
Metrics for Image *i* and Class Stain.

	Metrics for image *i*		Metrics for class (stain)	
Precision	$P_i = \frac{TP_i}{TP_i + FP_i}$	(1)	$P = \left(\frac{1}{N}\right) \sum_{i=1}^{N} P_i$	(2)
Recall	$R_i = \frac{TP_i}{TP_i + FN_i}$	(3)	$R = \left(\frac{1}{N}\right) \sum_{i=1}^{N} R_i$	(4)
F-score (specifically F1-score)	$F1_i = 2 \cdot \frac{P_i \cdot R_i}{P_i + R_i}$	(5)	$F1 = \left(\frac{1}{N}\right) \sum_{i=1}^{N} F1_i$	(6)
Intersection over Union (IoU)	$IoU_i = \frac{Mask_i \cap Prediction_i}{Mask_i \cup Prediction_i}$	(7)	$IoU = \left(\frac{1}{N}\right) \sum_{i=1}^{N} IoU_i$	(8)

reference mask not identified as part of the segmentation are counted as false negatives.

The Tukey's Honestly Significant Difference (HSD) test was conducted to assess the outcomes. This statistical method, applied in conjunction with analysis of variance (ANOVA), is designed to detect significant differences between mean values. A significance level of $p < 0.05$ was used. A single factor was employed to represent the combination of architectures, optimizers, and learning rates to facilitate the interpretation of these tests. Additionally, boxplots and overlays of predictions on images were generated and employed in the analysis, interpretation, and discussion of results. In addition, precision and recall were calculated on a per-image basis where relevant to the discussion.

4. Results

Table 4 and Appendices 01, 02, 03, and 04 detail the performance metrics of the tested models. The metrics of precision, recall, F-

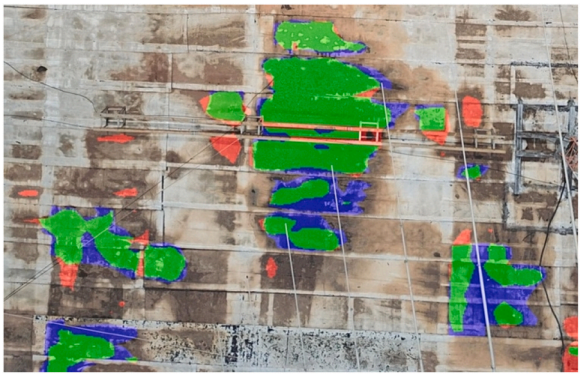

Fig. 9. Color representation for TP (green), FP (red), and FN (blue).

Table 4
Metric results: precision, recall, F-score, and IoU.

Metrics	Architectures	Optimizer	Learning rate	CLD	Mean (± sd) [%]	Median (IQR) [%]
Precision	FCN	Adagrad	1e−2	a	79,69 (± 3,06)	78,73 (4,32)
	DeepLabV3	Adam	1e−4	a	79,43 (± 5,97)	79,03 (10,24)
	SegFormer	Adam	1e−4	ab	69,35 (± 7,54)	69,25 (5,68)
Recall	FCN	Adagrad	1e−2	a	67,81 (± 6,27)	65,36 (6,77)
	DeepLabV3	Adam	1e−4	a	65,00 (± 8,82)	64,90 (10,09)
	SegFormer	Adam	1e−4	ab	59,61 (± 12,29)	62,88 (16,72)
F-score	FCN	Adagrad	1e−2	a	73,09 (± 3,53)	71,74 (6,00)
	DeepLabV3	Adam	1e−3	a	71,91 (± 3,03)	72,29 (4,50)
	SegFormer	Adam	1e−4	abcd	63,61 (± 9,60)	65,70 (13,70)
IoU	FCN	Adagrad	1e−2	a	57,70 (± 4,42)	55,93 (7,50)
	DeepLabV3	Adam	1e−3	ab	56,23 (± 3,67)	56,61 (5,50)
	SegFormer	Adam	1e−4	abcde	47,29 (± 10,25)	48,93 (14,40)

score, and IoU are grouped into combinations identified in the CLD column from "a" to "i" by the results of the Tukey HSD test. Fig. 10 presents the results in boxplot graphs, which illustrate the dispersion of values around the median based on the architectures, optimizers, and learning rates employed.

Upon examination of the tables, it was determined that the highest averages for all metrics were achieved by the FCN when optimized with Adagrad at a learning rate of 1e-2. In this case, the averages obtained for precision, recall, F-score, and IoU were 79.69 %, 67.81 %, 73.09 %, and 57.70 %, respectively. It is also noteworthy that several techniques employed yielded statistically equivalent averages across all metrics. This phenomenon was predominantly observed in the context of DeepLabV3 networks, though it did manifest to a lesser extent in the case of the SegFormer architecture. However, when evaluated using the most utilized optimization strategies, SegFormer demonstrated statistically inferior results across all four metrics. Furthermore, it can be observed that in some cases, the performance of the networks was zero or close to zero, which is likely due to optimization issues such as exploding gradients [34]. Notably, this phenomenon occurred with all neural networks optimized using SGD at a learning rate 1e-4.

5. Discussion

Figs. 11–17 depict the overlap of predictions generated by the FCN, DeepLabV3, and SegFormer networks optimized by different strategies. In conjunction with the metrics, our study suggests that the FCN and DeepLabV3 networks were successful in segmenting the darker stains, which are indicative of severity, from the lighter stains. This successful segmentation demonstrates the models' ability to learn fine-grained attributes for segmentation.

Given the complex visual characteristics of flat slab surfaces, especially when images are captured by UAVs, it is important to assess the robustness of the models used for stain segmentation. Our study found that the neural networks evaluated were able to effectively identify moisture stains, even in areas with drainage pipes, equipment, and debris. Notably, the FCN and DeepLabV3 networks successfully segmented dark stains near parapets and drains, despite the visual interference from surrounding objects. This performance can be observed in Figs. 11, 12, and 13, where the segmentations performed by the networks are clearly visible.

In particular, the FCN network, optimized with the Adagrad algorithm, demonstrated a solid ability to distinguish stains from other objects. At the same time, DeepLabV3 performed well in regions prone to water stagnation, places that often cause confusion and tend to hinder the work of other traditional algorithmic models [35].

However, the presence of false positives, evident around objects such as equipment and debris, supports the assumption of possible limitations in the networks' robustness. In some instances, debris's color and texture were incorrectly associated with stains. This suggests that, although the models demonstrate robustness in certain areas, improvements could be made, such as adjusting training

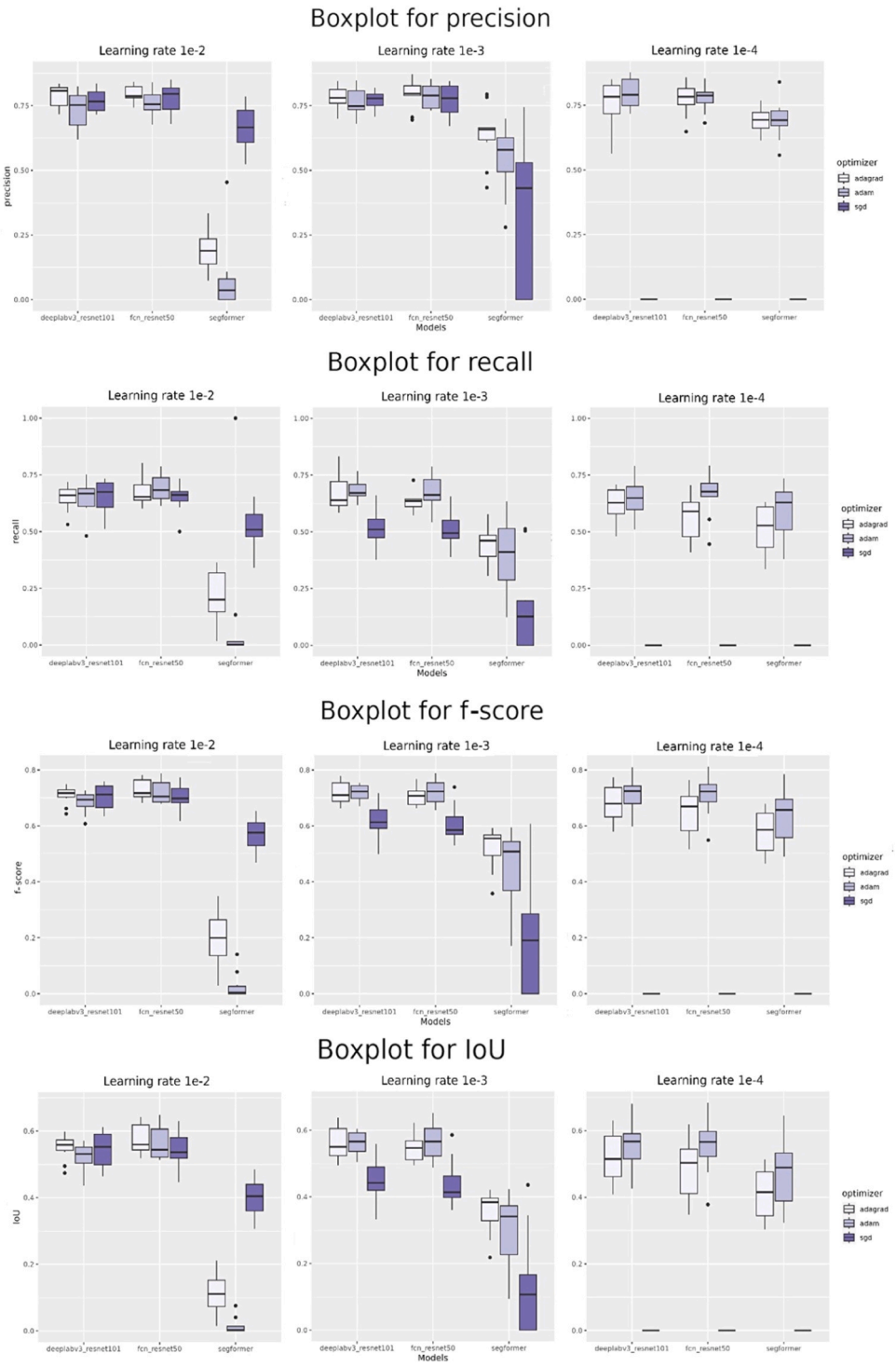

Fig. 10. Boxplot of the results.

techniques and enhancing data quality, for example, by using more sophisticated cameras and approaches such as transfer learning [36].

Regarding the metrics, F-measure and IoU were both helpful for evaluating the robustness of the model, providing a balance between precision and recall. Observing these metrics, we can better understand how the models deal with simple and complex tasks such as segmentation [37]. Especially in complex cases, where fine details or small objects were involved, the accuracy of the models

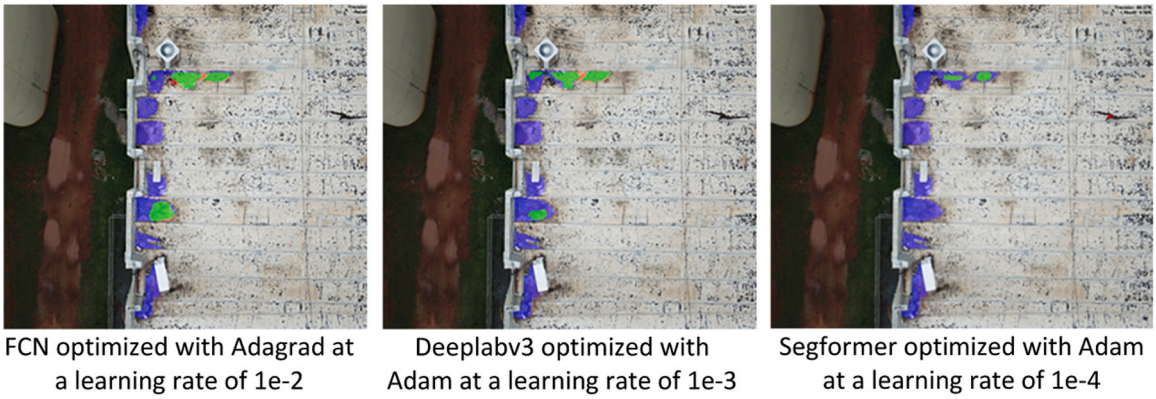

Fig. 11. Segmentation using FCN, Deeplabv3, and Segformer.

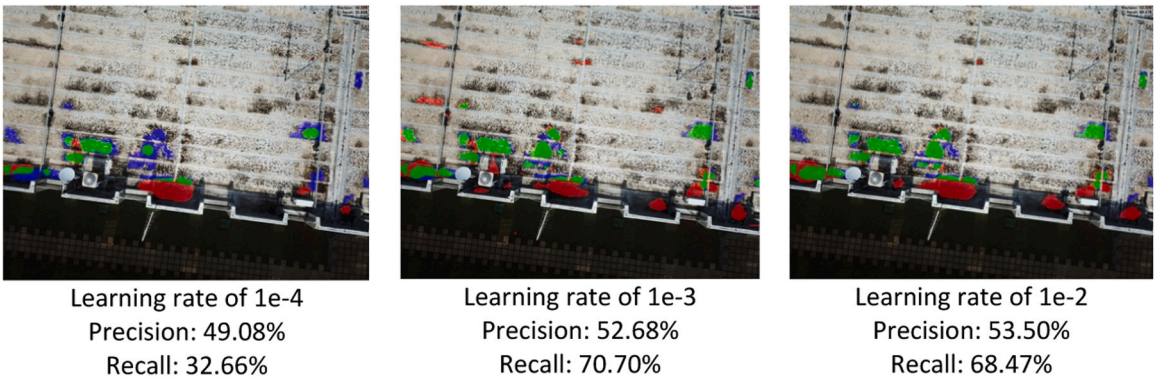

Fig. 12. Segmentation using FCN with Adagrad optimizer at different learning rates.

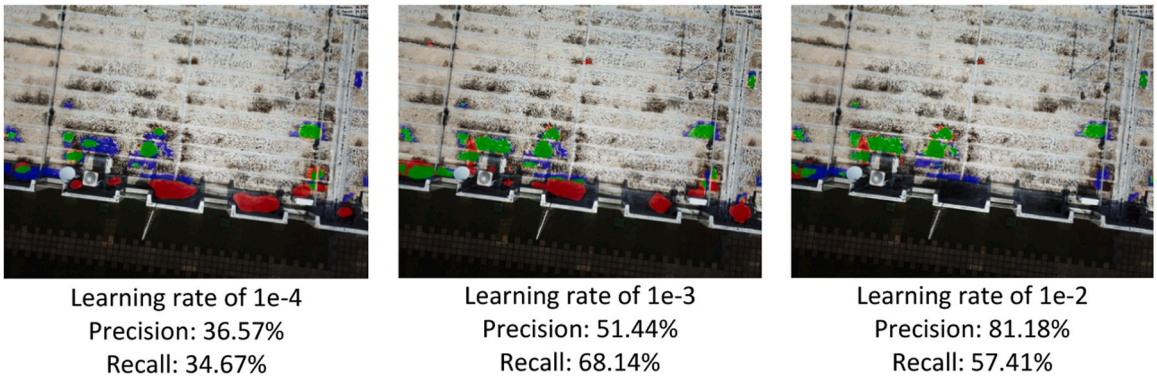

Fig. 13. Segmentation using DeepLabV3 with Adagrad optimizer at different learning rates.

tended to be lower, indicating room for improvement.

The results achieved by the SegFormer networks indicate that optimizing this architecture is a more challenging task, possibly since they are based on transformers rather than convolutions. Since the conception of the Vision Transformer (ViT) by Dosovitskiy et al. [38], it has been generally accepted that the performance of transformer-based neural networks can be better than that of CNNs, provided that the datasets used for training are larger. This phenomenon may be at play in the case of SegFormer, although the task to be performed is segmentation [39]. Therefore, this architecture could be re-evaluated in future work with larger datasets.

Figs. 12 and 13 show the overlap of predictions made by FCN and DeepLabV3 models for an image of the region near the parapet. In some cases, the errors made by the tested models are close to the drainage pipes, parapets, and equipment. These regions are more critical because water stagnation frequently occurs due to a lack of drain cleaning, drainage pipe blockages, and improperly installed

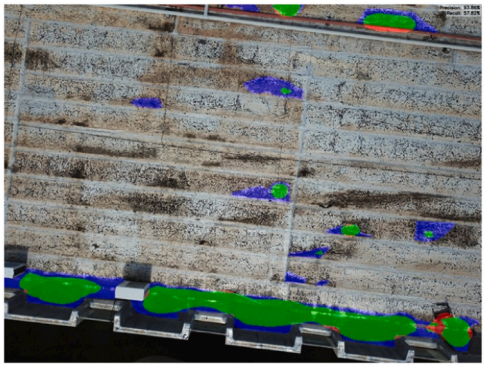

Fig. 14. Water stagnation close drainage pipes - DeepLabV3 optimized with Adagrad and a learning rate 1e-4.

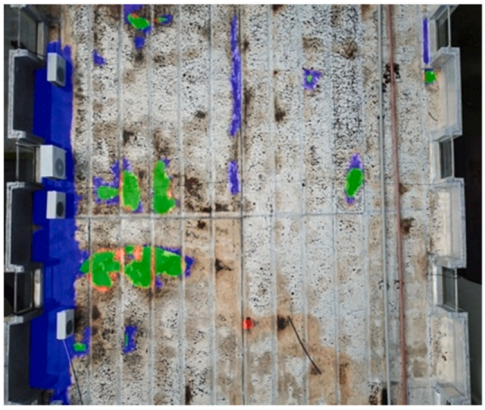

Fig. 15. Segmentation of false negatives (FN=blue) in stains close drainage pipes, equipment, and parapets - FCN optimized with Adagrad and a learning rate 1e-4.

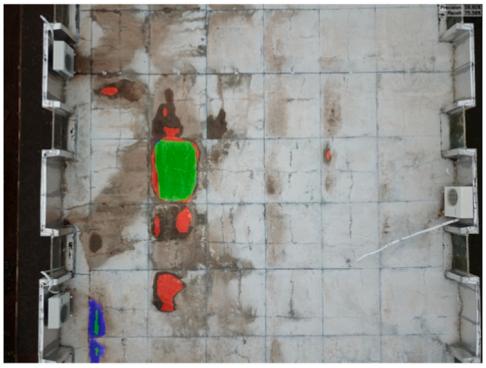

Fig. 16. Light and dark moisture stains on mechanical protection with mortar - FCN optimized with Adagrad and rate of 1e-4.

equipment.

Since the occurrence of stains may be more common in these regions, it is possible that the decisions made by the networks are being influenced by spatial information and visual characteristics of the parapets rather than by properties that are correctly indicative of stains, leading to a higher number of false positives. On the other hand, the presence of false negatives in the regions near the parapet in specific images indicates that, although the position relative to it may not be relevant, it might be influencing the segmentation. If this is the case, it is possible that collecting new data and using different data augmentation and regularization techniques could improve the models' performance.

It is important to consider the section regarding spatial resolution in drone cameras because object detection accuracy is highly dependent on photo quality for modeling [40]. A higher-resolution camera results in better object detection accuracy [13]. Another

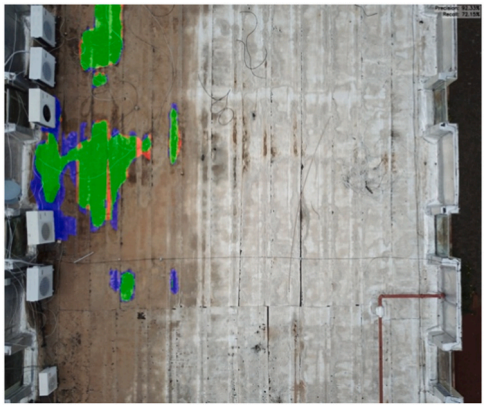

Fig. 17. Light and dark moisture stains on the aluminized blanket - DeepLabV3 optimized with Adagrad and rate of 1e-2.

important factor is the flight height for capturing higher-resolution images [41]. There are now cameras with higher-resolution sensors that could improve the results of mapping stains on flat roofs and other building detection applications.

The dataset used in the experiment contains images of both asphalt blankets with mechanical protection by a mortar layer (Fig. 16) and asphalt blankets with protection by an aluminized film (Fig. 17). Lighter or darker stains occur similarly on both types of materials. Over time, the aluminum film wears out without necessarily causing water pooling or moisture concentration. The networks learned well to identify the dark stains in both situations without significant confusion.

The orientation of the asphalt waterproofing strips affects water drainage and rubbish accumulation, as it directly influences how water flows to the drainage pipes. Installing the strips against the direction of water flow is inappropriate, resulting in more segmented stains parallel to the parapets (Fig. 18). In the image, it is even possible to identify speckling on the aluminized film. However, the networks' performance shows no distinction between the two technical protection solutions, and the speckling was not segmented. This demonstrates that the aluminum did not hinder detection, and the models' learning was as expected.

In some parts of the flat roof, rubbish resulting from maintenance actions was identified. Pieces of glass remaining from mortars and discarded condenser units were segmented by the networks as false positives (FP=red), possibly due to the color of the rubbish (Fig. 19).

In the research, the challenge is identifying stains related to the segmentation of objects with varying sizes, shapes, and colors in images of complex environments due to natural exposure conditions and various equipment attached to the surface under study. The presence of similar pixels in both light and dark stains, as well as rubbish on the flat roof surface, can confuse operators, complicating the effective training of the segmentation model. This difficulty is similar to the problem addressed by Tang et al. [42] in their study on landslides with seismic data, where the SegFormer and DeepLabV3 models faced similar obstacles, such as inadequate consideration of morphological features and difficulty in dealing with holes in the center of objects in the image.

Using the F-measure was important for balancing the precision and recall metrics, providing a clearer overview of the data resulting from the models. This facilitated a more integrated analysis of the results. Additionally, the Intersection over Union (IoU) metric provided a more detailed assessment of the location of the objects of interest identified by the models. These results reinforce the

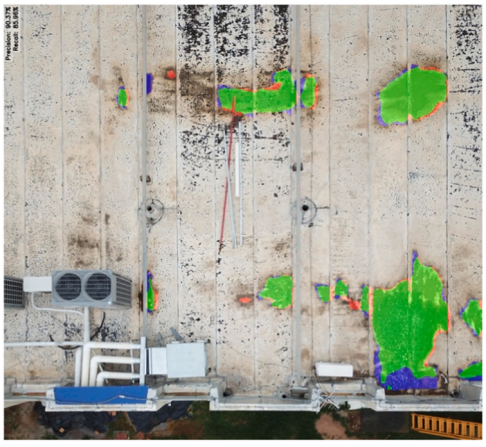

Fig. 18. Aluminized blankets applied in the direction of the slope for drainage pipe (the most suitable technique for rainwater drainage) - FCN with Adagrad optimizer and 1e-2.

Fig. 19. False positive segmentation of rubbish on the flat roof - FCN with Adagrad optimizer and a learning rate 1e-4.

hypothesis that the dataset is complex and exhibits considerable variation, contributing to the dissimilarities observed between the data.

6. Conclusion

In conclusion, the research tested and compared two convolutional neural networks for semantic segmentation (FCN with the ResNet50 backbone and DeepLabV3 with the ResNet101 backbone) and a deep artificial neural network of the transformer type (SegFormer with the MiT-B1 backbone). The best performance results were achieved by combining the FCN network with the Adagrad optimizer at a learning rate 1e-2, reaching a precision of 79.69 % and a recall of 67.81 %.

Despite the inherent complexity of stains on flat roofs, the relevant experiment carried out in this research yielded fully satisfactory and promising results. Even when dealing with features such as variability in shape, size, color, brightness, positioning, proximity to equipment, and interference from other elements in the same plane, the experiment showed that the networks could differentiate between dark and light stains. As indicators of more degraded points or regions on the flat roof, the segmentation of dark stains detects damage caused by moisture and dirt accumulation. Therefore, it is a relevant tool for building inspections in complex environments requiring preventive maintenance.

For future work, it is planned to expand the dataset to increase the robustness of the algorithms, as a more extensive and diverse dataset will help better handle different scenarios. Tools such as Class Activation Map (CAM) could help improve the models' ability to interpret situations and make decisions in the segmentation of stains by deep neural networks. Additionally, other annotation techniques and segmentation models could be considered to improve performance in detecting several types of stains on flat roofs.

Finally, the main driver of this research was the use of machine learning to predict the need for maintenance before critical problems arise in flat roofs. Predictive modeling for a maintenance management system presents itself as a challenging path for the continuity of experiments. The integrated approaches in developing continuous learning models, capable of adjusting their predictions as new data is received, show promise for the construction industry.

For this, the consolidation of a solid database of building inspection, maintenance management, and sensory and climate monitoring is necessary to feed Digital Twins in BIM systems as information management resources that enhance prediction accuracy and optimize the management of building degradation. The positive impact of emerging technologies in the building construction and maintenance sector stands out, aiming for greater safety, sustainability, and resource savings.

CRediT authorship contribution statement

Gabriel Toshio Hirokawa Higa: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Leonardo Rabero Lescano:** Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Lenildo Santos da Silva:** Formal analysis, Data curation. **Vanda Alice Garcia Zanoni:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Hemerson Pistori:** Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Cesar Ivan Alvarez:** Formal analysis. **Lara Monalisa Alves dos Santos:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has received financial support from the Dom Bosco Catholic University and the Foundation for the Support and Development of Education, Science and Technology from the State of Mato Grosso do Sul, FUNDECT. Some of the authors have been

awarded with Scholarships from the University of Brasília, Brazilian National Council of Technological and Scientific Development, CNPq and the Coordination for the Improvement of Higher Education Personnel, CAPES.

Appendice 01. - Results of precision metric

Architectures	Optimizer	Learning rate	CLD	mean (± sd) [%]	median (IQR) [%]
FCN	**Adagrad**	**1e−2**	**a**	**79.69 (± 3.06)**	**78.73 (4.32)**
DeepLabV3	**Adam**	**1e−4**	**a**	**79.43 (± 5.97)**	**79.03 (10.24)**
FCN	Adagrad	1e−3	a	79.43 (± 5.72)	79.62 (3.77)
FCN	Adam	1e−3	ab	78.69 (± 4.86)	78.93 (8.49)
DeepLabV3	Adagrad	1e−2	ab	78.61 (± 4.56)	80.77 (6.93)
FCN	Adam	1e−4	ab	77.79 (± 5.12)	78.76 (4.02)
FCN	SGD	1e−2	ab	77.74 (± 5.98)	79.55 (8.20)
DeepLabV3	Adagrad	1e−3	ab	77.73 (± 4.86)	78.02 (5.27)
FCN	Adagrad	1e−4	ab	77.39 (± 6.34)	78.31 (6.21)
DeepLabV3	SGD	1e−3	ab	77.04 (± 3.33)	77.81 (4.43)
DeepLabV3	SGD	1e−2	ab	77.04 (± 4.65)	76.60 (7.18)
FCN	SGD	1e−3	ab	76.87 (± 6.45)	77.86 (10.14)
DeepLabV3	Adam	1e−3	ab	76.48 (± 5.15)	74.80 (7.35)
DeepLabV3	Adagrad	1e−4	ab	75.99 (± 8.79)	78.30 (10.91)
FCN	Adam	1e−2	ab	75.62 (± 5.18)	75.52 (5.79)
DeepLabV3	Adam	1e−2	ab	73.51 (± 7.15)	75.21 (11.40)
SegFormer	**Adam**	**1e−4**	**ab**	**69.35 (± 7.54)**	**69.25 (5.68)**
SegFormer	Adagrad	1e−4	ab	69.08 (± 5.37)	69.32 (6.07)
SegFormer	SGD	1e−2	abc	66.16 (± 8.92)	66.60 (12.39)
SegFormer	Adagrad	1e−3	bc	64.01 (± 11.16)	65.79 (4.56)
SegFormer	Adam	1e−3	c	54.11 (± 13.04)	57.92 (13.18)
SegFormer	SGD	1e−3	d	32.71 (± 29.51)	43.09 (52.98)
SegFormer	Adagrad	1e−2	de	19.37 (± 8.51)	18.89 (9.71)
SegFormer	Adam	1e−2	ef	7.91 (± 13.77)	3.60 (8.00)
DeepLabV3	SGD	1e−4	f	0.00 (± 0.00)	0.00 (0.00)
FCN	SGD	1e−4	f	0,00 (± 0.00)	0.00 (0.00)
SegFormer	SGD	1e−4	f	0,00 (± 0.00)	0.00 (0.00)

APPENDICE 02. - Results of recall metric

Architectures	Optimizer	Learning rate	CLD	mean (± sd) [%]	median (IQR) [%]
FCN	**Adagrad**	**1e−2**	**a**	**67.81 (± 6.27)**	**65.36 (6.77)**
DeepLabV3	**Adam**	**1e−4**	**a**	**65.00 (± 8.82)**	**64.90 (10.09)**
FCN	Adagrad	1e−3	a	64.08 (± 5.19)	63.57 (3.29)
FCN	Adam	1e−3	ab	67.07 (± 7.77)	66.25 (9.12)
DeepLabV3	Adagrad	1e−2	ab	65.01 (± 5.82)	65.99 (5.87)
FCN	Adam	1e−4	ab	66.15 (± 9.82)	67.73 (5.84)
FCN	SGD	1e−2	ab	64.71 (± 6.22)	66.21 (4.30)
DeepLabV3	Adagrad	1e−3	ab	67.05 (± 8.03)	63.93 (10.49)
FCN	Adagrad	1e−4	ab	56.51 (± 10.18)	58.99 (15.12)
DeepLabV3	SGD	1e−3	ab	51.71 (± 8.07)	51.00 (8.17)
DeepLabV3	SGD	1e−2	ab	65.59 (± 7.21)	67.49 (10.84)
FCN	SGD	1e−3	ab	50.84 (± 7.96)	49.45 (7.93)
DeepLabV3	Adam	1e−3	ab	68.20 (± 4.77)	67.13 (5.01)
DeepLabV3	Adagrad	1e−4	ab	61.92 (± 7.42)	62.81 (10.54)
FCN	Adam	1e−2	ab	69.20 (± 6.07)	68.32 (9.19)
DeepLabV3	Adam	1e−2	ab	64.97 (± 7.75)	66.79 (7.82)
SegFormer	**Adam**	**1e−4**	**ab**	**59.61 (± 12.29)**	**62.88 (16.72)**
SegFormer	Adagrad	1e−4	ab	51.28 (± 10.95)	52.76 (17.84)
SegFormer	SGD	1e−2	abc	51.42 (± 9.45)	50.86 (9.84)
SegFormer	Adagrad	1e−3	bc	44.70 (± 8.38)	46.09 (9.36)
SegFormer	Adam	1e−3	c	39.38 (± 17.11)	41.09 (22.65)
SegFormer	SGD	1e−3	d	16.58 (± 19.90)	12.65 (19.59)
SegFormer	Adagrad	1e−2	de	21.39 (± 11.22)	20.02 (17.30)
SegFormer	Adam	1e−2	ef	11.64 (± 31.31)	00.25 (1.46)
DeepLabV3	SGD	1e−4	f	0.00 (± 0.00)	0.00 (0.00)
FCN	SGD	1e−4	f	0.00 (± 0.00)	0.00 (0.00)
SegFormer	SGD	1e−4	f	0.00 (± 0.00)	0.00 (0.00)

APPENDICE 03. - Results of F-score metric

Architectures	Optimizer	Learning rate	CLD	mean (± sd) [%]	median (IQR) [%]
FCN	**Adagrad**	**1e−2**	**a**	**73.09 (± 3.53)**	**71.74 (6.00)**
FCN	Adam	1e−2	a	72.06 (± 4.06)	70.46 (6.93)
FCN	Adam	1e−3	a	72.06 (± 4.49)	72.30 (6.75)
DeepLabV3	**Adam**	**1e−3**	**a**	**71.91 (± 3.03)**	**72.29 (4.50)**
DeepLabV3	Adagrad	1e−3	a	71.62 (± 4.12)	70.99 (6.63)
DeepLabV3	Adam	1e−4	a	71.29 (± 6.88)	72.42 (6.27)
FCN	Adam	1e−4	a	71.24 (± 7.57)	72.29 (6.22)
DeepLabV3	Adagrad	1e−2	a	70.90 (± 3.28)	71.70 (2.58)
FCN	Adagrad	1e−3	ab	70.72 (± 3.66)	70.71 (4.80)
DeepLabV3	SGD	1e−2	ab	70.56 (± 4.54)	71.17 (7.70)
FCN	SGD	1e−2	ab	70.33 (± 4.32)	69.83 (5.12)
DeepLabV3	Adam	1e−2	abc	68.37 (± 3.85)	69.38 (4.12)
DeepLabV3	Adagrad	1e−4	abc	67.93 (± 6.91)	67.98 (10.54)
FCN	Adagrad	1e−4	abc	65.03 (± 8.67)	66.97 (12.23)
SegFormer	**Adam**	**1e−4**	**abcd**	**63.61 (± 9.60)**	**65.70 (13.70)**
DeepLabV3	SGD	1e−3	abcd	61.64 (± 6.49)	61.29 (6.65)
FCN	SGD	1e−3	abcd	60.81 (± 6.42)	58.52 (6.27)
SegFormer	Adagrad	1e−4	bcd	58.19 (± 7.95)	58.60 (13.29)
SegFormer	SGD	1e−2	cde	56.95 (± 6.04)	57.57 (8.18)
SegFormer	Adagrad	1e−3	de	52.14 (± 7.70)	55.52 (7.33)
SegFormer	Adam	1e−3	e	44.62 (± 15.07)	50.79 (17.51)
SegFormer	SGD	1e−3	f	20.61 (± 22.25)	19.01 (28.52)
SegFormer	Adagrad	1e−2	f	19.73 (± 9.43)	19.92 (12.80)
SegFormer	Adam	1e−2	g	2.72 (± 4.67)	0.49 (2.65)
DeepLabV3	SGD	1e−4	g	0.00 (± 0.00)	0.00 (0.00)
FCN	SGD	1e−4	g	0.00 (± 0.00)	0.00 (0.00)
SegFormer	SGD	1e−4	g	0.00 (± 0.00)	0.00 (0.00)

APPENDICE 04. - Results of IoU metric

Architectures	Optimizer	Learning rate	CLD	mean (± sd) [%]	median (IQR) [%]
FCN	**Adagrad**	**1e−2**	**a**	**57.70 (± 4.42)**	**55.93 (7.50)**
FCN	Adam	1e−3	a	56.50 (± 5.51)	56.62 (8.27)
FCN	Adam	1e−2	a	56.48 (± 5.04)	54.40 (8.46)
DeepLabV3	**Adam**	**1e−3**	**ab**	**56.23 (± 3.67)**	**56.61 (5.50)**
DeepLabV3	Adagrad	1e−3	ab	55.93 (± 5.04)	55.03 (8.12)
FCN	Adam	1e−4	ab	55.79 (± 8.80)	56.61 (7.54)
DeepLabV3	Adam	1e−4	ab	55.78 (± 8.25)	56.77 (7.54)
DeepLabV3	Adagrad	1e−2	abc	55.01 (± 3.85)	55.88 (3.13)
FCN	Adagrad	1e−3	abc	54.82 (± 4.44)	54.69 (5.69)
DeepLabV3	SGD	1e−2	abc	54.68 (± 5.38)	55.26 (9.16)
FCN	SGD	1e−2	abc	54.39 (± 5.11)	53.65 (6.16)
DeepLabV3	Adam	1e−2	abcd	52.06 (± 4.35)	53.12 (4.81)
DeepLabV3	Adagrad	1e−4	abcd	51.81 (± 7.93)	51.50 (12.20)
FCN	Adagrad	1e−4	abcde	48.73 (± 9.48)	50.34 (13.32)
SegFormer	**Adam**	**1e−4**	**abcde**	**47.29 (± 10.25)**	**48.93 (14.40)**
DeepLabV3	SGD	1e−3	bcdef	44.83 (± 6.75)	44.19 (7.03)
FCN	SGD	1e−3	cdef	43.98 (± 6.95)	41.37 (6.44)
SegFormer	Adagrad	1e−4	def	41.43 (± 7.90)	41.52 (13.18)
SegFormer	SGD	1e−2	efg	40.03 (± 5.86)	40.42 (7.99)
SegFormer	Adagrad	1e−3	fg	35.58 (± 6.71)	38.43 (6.78)
SegFormer	Adam	1e−3	g	29.75 (± 11.81)	34.09 (14.68)
SegFormer	SGD	1e−3	h	13.20 (± 15.42)	10.65 (16.63)
SegFormer	Adagrad	1e−2	hi	11.22 (± 5.79)	11.06 (7.90)
SegFormer	Adam	1e−2	i	1.43 (± 2.50)	0.24 (1.34)
DeepLabV3	SGD	1e−4	i	0.00 (± 0.00)	0.00 (0.00)
FCN	SGD	1e−4	i	0.00 (± 0.00)	0.00 (0.00)
SegFormer	SGD	1e−4	i	0.00 (± 0.00)	0.00 (0.00)

Data availability

Data will be made available on request.

References

[1] I. Flores-Colen, J. Brito, V.P. Freitas, Stains in facades rendering diagnosis and maintenance techniques classification, Constr. Build. Mater. 22 (2008) 211–221, https://doi.org/10.1016/j.conbuildmat.2006.08.023.

[2] M. Torres-Gonzáles, J. Valença, B.O. Santos, A. Silva, M.P. Mendes, StainView: a fast and reliable method for mapping stains in facades using image classification in hsv and cielab colour space, Remote Sens. 15 (2023) 2895, https://doi.org/10.3390/rs15112895.

[3] L.M.A. Santos, V.A.G. Zanoni, E. Bedin, H. Pistori, Deep learning applied to equipment detection on flat roofs in images captured by UAV, Case Stud. Constr. Mater. 18 (2023) e01917, https://doi.org/10.1016/j.cscm.2023.e01917.

[4] J. González-Domínguez, G. Sánchez-Barroso, J.G. Sanz-Calcedo, Preventive maintenance optimisation of accessible flat roofs in healthcare centres using the Markov chain, J. Build. Eng. 32 (2020) 101775, https://doi.org/10.1016/j.jobe.2020.101775.

[5] L.B. Staffa Junior, D.B. Costa, J.L.T. Nogueira, A.S. Silva, Web platform for building roof maintenance inspection using UAS and artificial intelligence, Int. J. Build. Pathol. Adapt. (2023) 01–25, https://doi.org/10.1108/IJBPA-12-2022-0186.

[6] J. Xu, F. Zeng, W. Liu, T. Takahashi, Damage Detection and Level Classification of Roof Damage after Typhoon Faxai Based on Aerial Photos and Deep Learning, Appl. Sci. 12 (2022) 4912, https://doi.org/10.3390/app12104912.

[7] H. Perez, J.H.M. Tah, A. Mosavi, Deep Learning for Detecting Building Defects Using Convolutional Neural Networks, Sensors 19 (2019) 3556, https://doi.org/10.3390/s19163556.

[8] J. Valença, M.P. Mendes, N. Jouen, K. Morin, N. Olivo, A. Silva, Emotional Maintenance: A Digital Model to Support Maintenance Decisions in Buildings Coatings, Eng. Proc. 53 (2023) 1–4, https://doi.org/10.3390/IOCBD2023-16488.

[9] D. Yudin, A. Naumov, A. Dolzhenko, E. Patrakova, Software for roof defects recognition on aerial photographs, J. Phys.: Conf. Ser. 1015 (2018) 032152, https://doi.org/10.1088/1742-6596/1015/3/032152.

[10] H. Kim, J. Lee, E. Ahn, S. Cho, M. Shin, S.-H. Sim, Concrete crack identification using a UAV incorporating hybrid image processing, Sensors 17 (2017) 2052, https://doi.org/10.3390/s17092052.

[11] Y.-J. Cha, R. Ali, J. Lewis, O. Büyükɵztürk, Deep learning-based structural health monitoring, Autom. Constr. 161 (2024) 105328, https://doi.org/10.1016/j.autcon.2024.105328.

[12] M.-T. Cao, Drone-assisted segmentation of tile peeling on building façades using a deep learning model, J. Build. Eng. 80 (2023) 108063, https://doi.org/10.1016/j.jobe.2023.108063.

[13] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, E.T. Matson, Advances and Challenges in Drone Detection and Classification Techniques: a state-of-the-art review, Sensors 24 (2023) 125, https://doi.org/10.3390/s24010125.

[14] S. Katakol, B. Elbarashy, L. Herranz, J.V. Weijer, A.M. Lopez, Distributed Learning and Inference with Compressed Images, IEEE Trans. Image Process. (2021) 3069–3083, https://doi.org/10.1109/tip.2021.3058545.

[15] Y. Mo, Y. Wu, X. Yang, F. Liu, Y. Liao, Review the state-of-the-art technologies of semantic segmentation based on deep learning, Neurocomputing 493 (2022) 626–646, https://doi.org/10.1016/j.neucom.2022.01.005.

[16] D. Dais, İ.E. Bal, E. Smyrou, V. Sarhosis, Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning, Autom. Constr. 125 (2021) 103606, https://doi.org/10.1016/j.autcon.2021.103606.

[17] M. Midwinter, Z.A. Al-Sabbag, C.M. Yeum, Unsupervised Semantic defect with pose prior, Comput. -Aided Civ. Infrastruct. Eng. 38 (2023) 2455–2471, https://doi.org/10.1111/mice.13001.

[18] S. Hao, Y. Zhou, Y. Guo, A Brief Survey on Semantic Segmentation with Deep Learning, Neurocomputing 406 (2020) 302–321, https://doi.org/10.1016/j.neucom.2019.11.118.

[19] O. Alshboul, G. Almasabha, A. Shehadeh, K. Al-Shboul, A comparative study of LightGBM, XGBoost, and GEP models in shear strength management of SFRC-SBWS, Structures 61 (2024) 106009, https://doi.org/10.1016/j.istruc.2024.106009.

[20] A. Kamariotis, K. Tatsis, E. Chatzi, K. Goebel, D. Straub, A metric for assessing and optimizing data-driven prognostic algorithms for predictive maintenance, Reliab. Eng. Syst. Saf. 242 (2024) 109723, https://doi.org/10.1016/j.ress.2023.109723.

[21] T.D. Akinosho, L.O. Oyedele, M. Bilal, A.O. Ajayi, M.D. Delgado, O.O. Akinade, et al., Deep learning in the construction industry: A review of present status and future innovations, J. Build. Eng. 32 (2020) 101827, https://doi.org/10.1016/j.jobe.2020.101827.

[22] A.J. Sánchez-Garrido, I.J. Navarro, J. García, V. Yepes, A systematic literature review on modern methods of construction in building: An integrated approach using machine learning, J. Build. Eng. 73 (2023) 106725, https://doi.org/10.1016/j.jobe.2023.10672.

[23] O. Alshboul, G. Almasabha, K.F. Al-Shboul, A. Shehadeh, A comparative study of shear strength prediction models for SFRC deep beams without stirrups using Machine learning algorithms, Structures 55 (2023) 97–111, https://doi.org/10.1016/j.istruc.2023.06.026.

[24] A. Shehadeh, O. Alshboul, G. Almasabha, Slope displacement detection in construction: an automated management algorithm for disaster prevention, Expert Syst. Appl. 237 (2024) 121505, https://doi.org/10.1016/j.eswa.2023.121505.

[25] Y. Xu, Y. Zhou, P. Sekula, L. Ding, Machine learning in construction: from shallow to deep learning, Dev. Built Environ. 6 (2021) 100045, https://doi.org/10.1016/j.dibe.2021.100045.

[26] X. Liu, Comparison of Four Convolutional Neural Network-Based Algorithms for Sports Image Classification, Proc. 2023 Int. Conf. Data Sci., Adv. Algorithm Intell. Comput. (DAI 2023), Adv. Intell. Syst. Res. (2024) 178–186, https://doi.org/10.2991/978-94-6463-370-2_20.

[27] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Conf. Comput. Vis. Pattern Recognit. 2 (2015) 3431–3441, https://doi.org/10.1109/CVPR.2015.7298965.

[28] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking Atrous Convolution for Semantic Image Segmentation, arXiv:1706.05587v3 [cs.CV], Comput. Vis. Pattern Recognit. 3 (2017) 1–14, https://doi.org/10.48550/arXiv.1706.05587v3.

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, et al., Attention Is All You Need, arXiv:1706.03762 [cs.CL], Comput. Lang. 7 (2023) 1–15, https://doi.org/10.48550/arXiv.1706.03762v7.

[30] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, SegFormer: Simple and efficient design for semantic segmentation with transformers, arXiv: 2105.15203 [cs.CV], Adv. Neural Inf. Process. Syst. 1 (2021) 1–18, https://doi.org/10.48550/arXiv.2105.15203.

[31] S. Ibrahim, N.A. Wahab, Optimizing neural network algorithms for submerged membrane bioreactor: A comparative study of OVAT and RSM hyperparameter optimization techniques, Water Sci. Technol. 89 (2024) 1701–1724, https://doi.org/10.2166/wst.2024.099.

[32] DÀ-JIÀNG Innovations Science and Technology - DJI. DJI Fly. Available at: ⟨https://www.dji.com/br/downloads/djiapp/dji-fly⟩. Accessed on: Dezember 8, 2023.

[33] ROBOFLOW (2023). Give your software the sense of sight. Available at: ⟨https://roboflow.com/⟩. Accessed on: Dezember 8, https://doi.org/10.1016/j.neucom.2022.01.005.

[34] X. Glorot, Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. International Conference on Artificial Intelligence and Statistics, 13th AISTATS. Proceedings…, 9 of JMLR: W&CP 9. Italy, 2010. Available at: ⟨https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf⟩. Accessed on: May 15, 2024.

[35] K.K. Brar, B. Goyal, A. Dogra, M.A. Mustafa, R. Majumdar, A. Alkhayyatf, V. Kukreja, Image segmentation review: Theoretical background and recent advances, Inf. Fusion 114 (2025) 102608, https://doi.org/10.1016/j.inffus.2024.102608.

[36] K. Xing, J. Ku, J. Zhao, A Novel Approach to Optimizing Convolutional Neural Networks for Improved Digital Image Segmentation, Int. J. Intell. Syst. 1 (2024) 10, https://doi.org/10.1155/2024/4337255.

[37] O. Rainio, J. Teuho, R. Klén, Evaluation metrics and statistical tests for machine learning, Sci. Rep. 14 (2024) 6086, https://doi.org/10.1038/s41598-024-56706-x.

[38] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., An image is worth 16x16 words: Transformers for image recognition at scale. ICLR 2021 arXiv:2010.11929v2 [cs.CV], (2021), 01-21, https://doi.org/10.48550/arXiv.2010.11929.

[39] H. Thisanke, C. Deshan, K. Chamith, S. Seneviratne, R. Vidanaarachchi, D. Herath, Semantic segmentation using Vision Transformers: A survey, Eng. Appl. Artif. Intell. 126 (2023) 106669, https://doi.org/10.1016/j.engappai.2023.106669.
[40] C.I. Alvarez-Mendoza, D. Guzman, J. Casas, M. Bastidas, J. Polanco, M. Valencia-Ortiz, et al., Predictive Modeling of Above-Ground Biomass in Brachiaria Pastures from Satellite and UAV Imagery Using Machine Learning Approaches, Remote Sens. 14 (2022) 5870, https://doi.org/10.3390/rs14225870.
[41] Y. Guo, J. He, J. Huang, Y. Jing, S. Xu, L. Wang, S. Li, G. Zheng, Effects of the Spatial Resolution of UAV Images on the Prediction and Transferability of Nitrogen Content Model for Winter Wheat, Drones 6 (2022) 299, https://doi.org/10.3390/drones6100299.
[42] X. Tang, Z. Tu, Y. Wang, M. Liu, D. Li, X. Fan, Automatic detection of coseismic landslides using a new transformer method, Remote Sens. 14 (2022) 2884, https://doi.org/10.3390/rs14122884.