Received January 7, 2021, accepted January 21, 2021, date of publication January 25, 2021, date of current version February 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3054468

Automated Monitoring of Construction Sites of Electric Power Substations Using Deep Learning

BRUNO ALBERTO SOARES OLIVEIRA[1], ABÍLIO PEREIRA DE FARIA NETO[2], ROBERTO MÁRCIO ARRUDA FERNANDINO[2], ROGÉRIO FERNANDES CARVALHO[2], AMANDA LOPES FERNANDES[3], AND FREDERICO GADELHA GUIMARÃES[1], (Senior Member, IEEE)

[1]Graduate Program in Electrical Engineering, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil
[2]SVA Tech, Belo Horizonte 30130-165, Brazil
[3]CPFL Energia, Campinas 13088-900, Brazil

Corresponding author: Frederico Gadelha Guimarães (fredericoguimaraes@ufmg.br)

This work was supported by the Research and Development Program of ANEEL by means of CPFL Energia and partnership with the company SVA Tech under Grant PD-00063-3073/2020. The work of Frederico Gadelha Guimarães was supported by CNPq, Brazil, under Grant 312991/2020-7.

ABSTRACT With each passing year, the consumption of electric energy in Brazil and the world increases, making it necessary to adopt measures such as the construction of new plants and the installation of power distribution structures. Monitoring for construction management in companies is still done in person and manually, resulting in expenses that could be avoided. That said, there are opportunities to automate such processes using artificial intelligence and, therefore, the main objective of this work is the development of an automated constructions management system, whose goal is to increase the management and monitoring of substation constructions with the remote monitoring. The system incorporates resources of deep learning to classify the components in bays, comparing the data generated in this recognition with the engineering projects to verify the progress of the installation of these components and generating indicators of conformity and evolution of the construction. To achieve the main objective, a comparison was made among four convolutional neural network architectures: DenseNet, Inception, ResNet, and SqueezeNet, in the classification task. The models were trained with thousands of images extracted from photos of different bays captured in the field and, additionally, data augmentation techniques were applied. The models were trained using transfer learning and fine tuning starting from pre-trained weights in the ImageNet data set. All models obtained results close to 100% in the images of the test set, hence it is possible to conclude that, for the proposed problem, there was no significant difference between the assertiveness of the architectures. The chosen model was part of the final application that monitors the construction management of the bays in the electricity substations.

INDEX TERMS Computer vision, computerized monitoring, construction management, image classification, machine learning.

I. INTRODUCTION

It is believed that in the coming decades the world will consume much more energy than today, after all, whenever there is access and availability of reliable energy, people will increasingly enjoy this good that today has become indispensable. Unfortunately, there is a huge portion of the world population that still does not have access to basic energy services. Also, the climate changes that have been occurring in recent years contribute directly to the development of alternative renewable energy solutions, which generates an increasing demand for new ventures in the electricity sector [1]. All this need gave rise to the demand to present technological innovations that help and contribute efficiently to supply the human needs related to the electric energy sector and, in parallel, contribute in some way to a better use of environmental resources.

The aid of technology for automation and optimization of problems in the industry has now become indispensable. One of the possible tools that can be used to solve these problems is machine learning methods, which are algorithms that analyze data, learn from it, and apply what they have learned to make decisions autonomously [2]. Machine learning can be found in several applications in the most varied areas, such

The associate editor coordinating the review of this manuscript and approving it for publication was Arif Ur Rahman.

as medicine, agriculture, engineering, in the electric power sector, telecommunications, among others [3]–[5].

The large volume of projects carried out by energy companies in Brazil increasingly demand human resources to monitor construction sites. For instance, in one company[1] each year the engineering team manages more than 85 constructions, of which approximately 20% (around 15 constructions) are for installing 15kV bays, which has an average cost of US$ 60,000 per construction. The bays are sets of elements with specific functions in the electrical system, which allow the composition of the substation in modules.

Due to the standardization of projects, the activities carried out become recurrent and procedural, creating opportunities for the use of technological resources for the automation of the management process, and, therefore, productivity gains for the employed human resources.

The development of projects in this sense makes possible a series of socio-environmental contributions, impacting on the safety and quality of life of the community, since the control of assets can prevent the interruption of power supply through the identification of possible system failures, or invasions that can cause damage to the structure, through data analysis [6], [7]. They enable greater agility in the construction of substations, facilitating the monitoring of managers to monitor the progress of the constructions, in addition to enabling simultaneous management of multiple constructions. This brings about an increase in the number of constructions being done in parallel, directly reflecting in the increase in the number of jobs offered.

Contributions and economic impacts can also be achieved, such as the gain resulting from changes in the company's administrative processes, in the costs of internal labor, materials, displacements, inputs, and time of execution of activities. The possibility of improving the management of assets, from the monitoring of the constructions through images to preventing the erroneous installation facilities during the execution of the construction. The need to increase the number of constructions in substations can be a bottleneck, since the number of people available and trained to accompaniment does not grow in the same proportion, hence it is necessary to create solutions for the accompaniment of constructions facilitating management [8]–[10].

That said, the main objective of this work is the development of an automated system for construction sites management, whose proposal is to increase the management and monitoring of constructions in electric power substations with remote monitoring. The system incorporates the resources of deep learning for the classification of components in bays, comparing the data generated in this recognition with the engineering projects, aiming to verify the progress of the installation of these components and generating indicators of conformity and evolution of the construction. To achieve the main objective, a comparison was made among different convolutional neural network architectures, implementing machine learning models that are capable of classifying the main elements that make up the bays in substations, aiming, from this model trained with a new data set, to classify such elements through a system that will serve to optimize the constructions management activities in substations.

The rest of the work is organized as follows: Section II presents the problem of interest addressed in the research, highlighting the difficulties of managers incorrectly monitoring the construction of bays in electrical substations. Section III addresses the theoretical framework, bringing the fundamental concepts to understand the research and related works. The methodology is presented in Section IV, showing each of the steps followed in the development of the research. Section V presents the results achieved by the proposed system. Finally, conclusions and future work are provided in Section VI.

II. PROBLEM OF INTEREST

Currently, the monitoring of activities of the constructions in the distribution substations has been carried out in person and manually, being susceptible to technical errors or even the bad faith of the employee. Regularly, in short periods defined by the company, those responsible must travel to each substation to carry out the process of monitoring and advancing the constructions.

In times of pandemic caused by COVID-19, the monitoring by managers who carry out construction management activities was completely out of date. The public authorities of each municipality regulate through sanitary barriers everyone who enters and leaves the city, causing a considerable delay in the managers' travel time. Also, given the importance of social isolation and the need to not have contact with other people, the employee responsible for carrying out this monitoring may be contributing to the spread of Sars-CoV-2, given this frequent displacement between different municipalities.

The main components of a substation bay are: panel, current transformers, power transformers, disconnect switches, and isolators. These bays are essential for the supply of electricity to society and, as the demand for energy grows, it becomes necessary to install more of this equipment. As the number of construction sites to be monitored increases, and considering the distance between them, it is more difficult for managers to track the progress of several projects at the same time. Fig. 1 shows two examples of bays.

It is possible to observe in Fig. 1 the disposition of each component in the implantation of a bay, which always occurs in an already defined sequence: first, the concrete support for the panel is built, installing the panel immediately on it. Subsequently, current transformers (C.T.'s) are added, with one pair for each phase. The next component to be installed are the power transformers (P.T.'s), which coupled to them there is a protection grid, popularly called ''anti birds'', also counting one for each phase. Then, the rest of the structure is assembled.

[1] Internal data from CPFL Energy, Brazil. CPFL stands for *Companhia Paulista de Força e Luz*.

FIGURE 1. Example of two bays.

Due to this standardization of projects, it is possible to develop a system that can assist in the management of constructions in electric power substations. Fig. 2 presents an overview of how the system is designed.

To enable the use of computer vision in the recognition and classification of objects with the characteristics of the components of a bay in a substation, the adoption of technology of artificial neural networks with deep learning is adequate. This technology makes intensive use of processing capacity to achieve the right degree of assertiveness. To improve the relationship between assertiveness and processing capacity, one of the intervention variables is the architecture of the convolutional neural network used, which was one of the specific objectives of the development of the system.

The developed platform aims to automate constructions of bays in substations and can be used and/or expanded to other energy companies that carry out constructions, so that they are able to increase the number of constructions in progress simultaneously guaranteeing the quality of the execution of the constructions, in addition, to improve activity management and schedule control.

III. THEORETICAL REFERENCE

A. DEEP LEARNING

Deep learning involves a class of machine learning algorithms whose main objective is to extract resources so that the computer can learn from examples and, later, make decisions for a given problem [11], [12].

In deep learning, a model learns to perform image, text, or sound classification tasks. Such deep learning models can often achieve very accurate results, sometimes exceeding the performance of human beings themselves.

Although deep learning was theorized in the 1980s [13], it has only been disseminated in recent years for two reasons: a large amount of data currently available and the advancement of research in the hardware area, increasing existing computational resources [14], [15].

Computer vision and deep learning can be used in various tasks, such as image and video recognition, image analysis and classification, visual media recreation, among other things [16]. Advances in computer vision [17], [18] and deep learning have evolved, mainly through Convolutional Neural Networks (CNNs).

B. CONVOLUTIONAL NEURAL NETWORKS

A convolutional neural network, also called ConvNet, is a machine learning algorithm that receives an image as input and, from it, can extract resources by assigning importance (weights and biases learned) to various aspects, thus being able to differentiate objects contained there [19], [20]. The structure of CNN was inspired by the organization of the

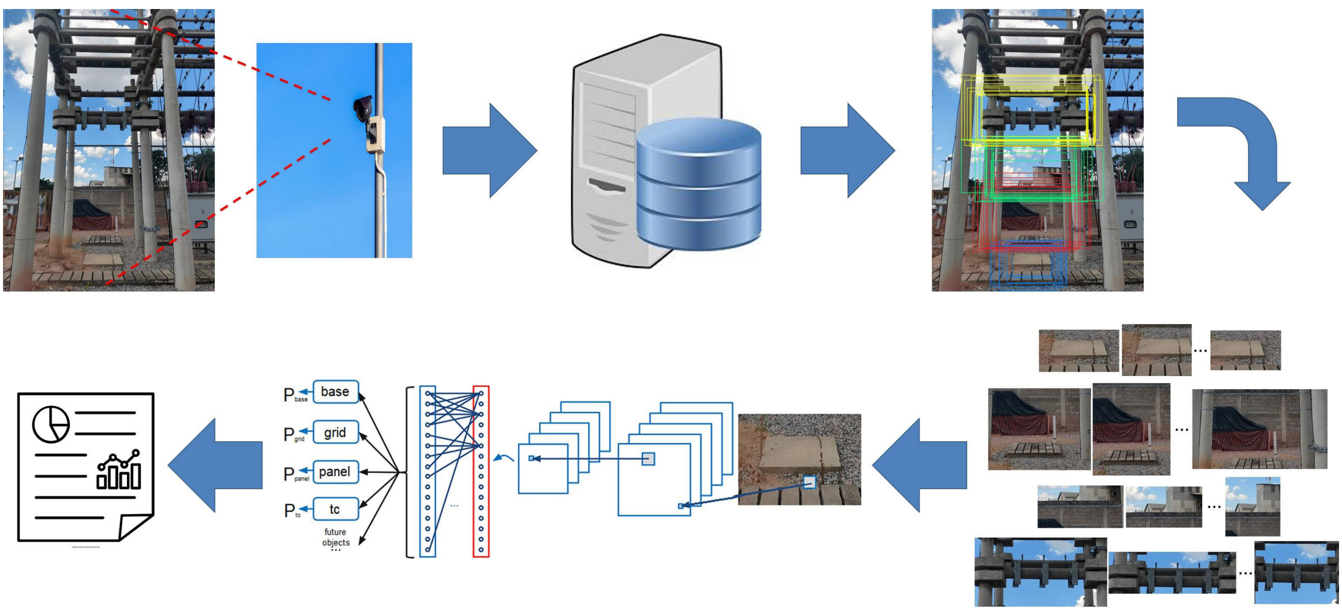

FIGURE 2. Overview of the functioning of the proposed system.

visual cortex of human beings. Individual neurons respond to stimuli that correspond to a specific restricted region of the visual field, known as the receptive field. The grouping of these fields overlaps, being able to cover the entire visual area [21].

Unlike conventional digital image processing techniques that use previously established filters, CNNs can automatically learn the best parameters of resource extractors. With this, they can obtain an extremely satisfactory performance in tasks of classification, detection, and segmentation of objects, surpassing the precision achieved by conventional techniques of digital image processing.

A CNN architecture is made up of convolution layers, pooling layers, and fully connected layers. Each convolution layer has a number of filters of certain sizes, which are responsible for extracting the high-level features from the input image. The pooling layers have the function of reducing the spatial size of the features involved, implying a reduction in the dimensionality of the data and consequently reducing the computational resource used for processing. Finally, the fully connected layers learn the nonlinear combinations of the features, making it possible, through an activation function at the end of the network, to classify objects [22].

There is a multitude of possibilities for the layering of the layers that make up a CNN. The deeper the architecture of a CNN, the greater the computational processing cost used, that is, the defined architecture varies according to the problem. Currently, some CNN architectures are available that have been pre-trained on large sets of images and have achieved quite satisfactory performance. Some of these well-known CNN architectures are DenseNet, Inception, ResNet, and SqueezeNet.

1) DENSENET

The densely connected CNN emerged from the work of [23], in which a network was proposed that, besides being deep, is also precise and efficient. The authors believed that if the connections between the layers were shorter, the training of the network would be better adjusted. Thus, the implementation of the architecture was defined with the premise that for each layer, with the feature maps obtained by the previous layers, these would also be used as input for all other subsequent layers.

The DenseNet architecture has some advantages over competing architectures: (1) the network can minimize the problem of gradient leakage, (2) it strengthens the propagation of features obtained through convolution operators, (3) it reuses such features, and (4) substantially reduces the number of network parameters. As presented by [23], the DenseNet architecture was evaluated in four object classification benchmark tasks (CIFAR10, CIFAR-100, SVHN, and ImageNet), obtaining significant improvements on the state of the art in large part of them, requiring less memory and computational performance.

2) INCEPTION

The CNN Inception architecture was first presented in the work of [24]. The authors agreed that the deeper the network and the greater the computational processing power, the greater the precision gains in the vast majority of tasks. However, they argued that computational efficiency and the limited number of parameters are still configurations that could be adjusted and improved, especially when working with a large data set and with digital image processing techniques.

That said, in the architecture proposed by [24], a means of using computational features in the most efficient way possible was explored, through appropriately factored convolutions and aggressive regularization. The CNN Inception architecture was tested in the ILSVRC 2012 classification challenge, obtaining a 21.2% error in the competition's top-1, surpassing the state of the art at the time.

3) RESNET

The residual neural network proposed in the work of [25] presents a CNN architecture structured on residual learning, facilitating the training of deep networks. The authors explicitly reformulated the learning layers using residual functions at the input of the layers, instead of using disconnected activation functions. With that, it was possible to show that the CNN ResNet architecture is easier to be optimized and can obtain high precision even with a considerably greater depth.

ResNet was evaluated in the ImageNet data set, obtaining an error of 3.57% in the test set, winning the first place ILSVRC 2015 for the object classification task. Currently, it is considered one of the main architectures of CNNs, being used in problems of classification and detection of objects, both in images and in videos.

4) SQUEEZENET

The vast majority of proposed CNN architectures focus on mainly improving accuracy. That said, in the work of [26] a CNN architecture was proposed that is much less deep than the competitors, called SqueezeNet. With equivalent accuracy, SqueezeNet offers at least three advantages over other architectures: (1) smaller networks require less communication between servers during training using distributed processing; (2) lean networks require less bandwidth to export the model trained in applications that require this feature; (3) smaller networks are more feasible to be used in applications that operate on embedded devices and hardware that have little memory.

The CNN SqueezeNet architecture achieves precision with a value close to that achieved by competing architectures in the set of ImageNet images and, in parallel, it still has about 50× fewer parameters than AlexNet, another CNN architecture. Also, with the techniques used to compress the model, SqueezeNet can be compressed so that it occupies only 0.5MB of disk space, about 510× less than AlexNet.

C. RELATED WORKS

No research was found that used deep learning and CNN to specifically solve the monitoring problem for managing constructions in electric power substations. Some works apply deep learning to detect faults in substations in general, and some others that classify objects present in given civil construction.

In the research developed by [27], it was argued that the substantial increase in infrared images in the electrical system of substations presents a new challenge, exemplifying how the assessment of the status of devices is traditionally done. To overcome this problem, the work of [27] proposes a method of automatic analysis of infrared images. The algorithm implemented first segments images into superpixels and then adopts CNN to classify objects. Good results have been achieved compared to other unsupervised training methods tested.

The use of deep learning and the Yolo object detector can be found in the research by [28], in which the authors applied the method for intelligent detection and recognition of high voltage distribution frames. The proposed system aims to perform some tasks, such as identifying the placement of a key, as well as recognizing and determining its state.

In the work of [29] a new approach to state recognition based on CNN of commutators in substations was proposed. Transfer learning was used in the model, in which the authors used the set of images ILSVRC2012, retraining the model with their images of the studied substation. In their experiments, they obtained quite satisfactory results and concluded that the proposed approach can be applied to the analyzed substation, which may reduce the cost of operation if the industry decides to implement the system.

The use of personal protective equipment (PPE) and collective protective equipment (CPE) is an important safety measure that aims to protect employees in the work environment. Unfortunately, due to lack of responsibility, there are still workers who persist in not using such equipment subjecting themselves to a greater risk of fatal accidents. To automate the process of detecting such employees, in the article of [30] the fastest R-CNN object detector was applied, to detect workers who do not have PPE. The experimental results showed that the method obtained high precision in detecting people who are going against the company's security policies.

For the proper management of constructions and the revisions of the plan during construction, it is necessary to understand the status of the construction in real-time. Aiming to assist in the development of solutions for civil construction, [31] proposes a deep learning method that aims to accurately recognize construction equipment. Due to the lack of available images of the objects covered, transfer learning was used, which classifies five classes: dump truck, excavator, loader, concrete mixer truck, and road roller. The results reached a value of 96.33% for the mean average precision (mAP). The proposed method can be used to infer the context of civil construction operations, generating data such as the progress of the work, productivity, and safety.

Similarly, [32] developed a system for monitoring construction management in civil construction, performing automatic detection of workers and excavators at a given construction site. The fastest R-CNN method was used, which obtained very satisfactory results, detecting the presence of workers and excavators with a high level of precision (91% and 95%). According to the authors, the accuracy of the proposed deep learning method exceeds that of current methods in detecting objects at construction sites.

In [33], deep learning was used for the detection of construction equipment, starting from a vacant subdivision until the work was completed. The analysis of the results confirms the superior performance in real-time of the proposed solution with an accuracy rate above 90%. The present study validates the practicality of object detection solutions based on deep learning for construction scenarios. Also, the solution can be used for various purposes, such as security monitoring, productivity assessments, and management decisions.

Part of Industry 4.0 applications is focused on developing autonomous smart substations. Among these, several methods of automatic meter reading were proposed with the invention of inspection robots. However, most solutions have difficulty in capturing quality meter images. In [34], a system based on deep learning and computer vision was proposed to obtain images of meters in acceptable conditions for future analysis. For this, the object detection method based on regions (Faster R-CNN) was used to find the exact position of the meter and then, the system adjusts the camera. The experimental results verify the stability and accuracy of recognition system which is proved to work well under different conditions.

In the work of [35], it was stated that the increase in the internal temperature of electrical instruments in power substations can cause unusual disturbances and damage to equipment. Therefore, as preventive measures for the problem mentioned, the authors proposed a new approach for the analysis of defects in this equipment, using infrared images and deep learning. The images were captured and their characteristics were extracted using a pre-trained CNN AlexNet. Then, the Random Forest and Support Vector Machine algorithms were used to classify whether the equipment had a defect or not. In an experimental analysis, the authors obtained an accuracy of 96% which, according to them, surpassed all other comparative approaches that used deep learning and others that used alternative technologies.

In recent years, several solutions have emerged for the monitoring of electric power substations, most of them based on computer vision and digital image processing algorithms. In the article by [36], the authors presented a proposal for the classification of ceramic and glass insulators by means of images. The Speeded-Up Robust Features (SURF) method was used to describe the image resources, and then the k-Nearest Neighbors (KNN) algorithm was applied to classify the obtained resources. The authors conducted experiments on insulators to verify the superiority of the technique.

The proposed method can be used in security, surveillance, and inspection system.

As previously mentioned, there are no studies in the current literature that use deep learning techniques in the classification of components in substation bays. The current research is about the development of an automated construction management system, in which a comparative study was made for different CNN architectures in a set of real images, collected in the field. The study carried out as part of the application that will assist the management of automated constructions for the construction of bays in electric power substations.

IV. METHODOLOGY

Each step of the execution of this work can be observed in a general way below.

1) Collection of images in the field.
2) Preparing the image set.
3) Division of images in training, validation, and testing, randomly varying into five different sets.
4) Application of data augmentation techniques to images.
5) Implementation of classifiers: DenseNet, Inception, ResNet, and SqueezeNet.
6) Conducting experiments to define the best CNN architecture.
7) Analysis of results and choice of the classifier.
8) Application and testing with the proposed system.

A. DATA COLLECTION AND PREPARATION

For the development of this research, the authors moved to different substations of the company CPFL Energia and collected, using a camera, 32 photos from different bays. Also, thinking of a set of images as variable as possible, the authors took care to capture the images at different times of the day, so that the dataset had images of bays with variation in the position of the sun reflecting the object, besides to different lighting conditions and climate. Subsequently, each of the components that are part of a bay was cut manually, obtaining a total of 23 images of the base class, and for the grid, panel, and C.T. objects, 29, 32, and 32 images, respectively. With that in mind, considering the model's generalization capacity, five different sets of images were generated, and, for each set, the 32 images were randomly divided into training, validation, and test set. The random division was performed as follows: out of a total of 100% of the amount for each of the elements, approximately 75% went to the training set and approximately 25% to the test set. Of the 100% of the training set after the first division, 80% remained and the other 20% went to the validation set. An example of this division is shown in Table 1, for the set of images that was part of the first experiment.

Table 1 exemplifies the randomness imposed in the division of training, validation, and test data, for the first set of images generated. The same randomness was also applied to the other four sets created. The separation of the original images, previously performed before applying the data augmentation operations, is justified by the fact that the models should not be allowed to be overfit during the training stage, because, if this occurs, the model may not be able to generalize well to the test set.

TABLE 1. Example of Randomness of the Set of Images.

IMAGE SET 1	
Training	
base	[1, 5, 8, 10, 11, 12, 13, 16, 17, 18, 21, 22, 23, 24, 25, 28, 30]
grid	[2, 3, 4, 5, 6, 7, 9, 12, 13, 17, 18, 19, 22, 23, 24, 25, 26, 27, 28, 31]
panel	[2, 4, 5, 6, 8, 9, 10, 11, 12, 15, 17, 19, 20, 22, 24, 25, 26, 28, 31, 32]
c.t.	[1, 2, 4, 6, 7, 8, 10, 11, 12, 14, 15, 16, 20, 21, 22, 28, 29, 30, 31, 32]
Validation	
base	[14, 19, 26]
grid	[1, 21, 29, 32]
panel	[1, 13, 14, 16, 18]
c.t.	[5, 13, 18, 19, 25]
Test	
base	[15, 20, 27]
grid	[10, 14, 15, 16, 20]
panel	[3, 7, 21, 23, 27, 29, 30]
c.t.	[3, 9, 17, 23, 24, 26, 27]

To increase the data, two operations were considered: cutting and horizontal inversion. It is important to note that such data augmentation operations were applied to all images, with the aim of having a sufficient number of images that can evaluate each of the trained models. However, as previously mentioned, before the application of these operations, each of the elements was properly separated into training, validation, and test sets, before the application of the augmentation operations. Thus, care was taken for the same element not to participate in the training stage of the model and also in the validation stage.

Manually, for each of the 32 captured photos, 20 images were generated for each of the four classes of objects, using randomly sized cutouts, but always keeping the object belonging to a particular class in question on the inside of the cut. Also, for each of the 20 cutouts, the horizontal inversion operation was used in each copy of the cut, having at the end of the application of the data augmentation 40 images of each object from each of the original 32 images, **totaling $32 \times 40 \times 4 = 5120$ images**. Fig. 3 illustrates how each cut was manually selected in each of the 32 photos captured in the field, so that subsequently these objects form part of the set of images for CNN training.

The yellow rectangles contained in Fig. 3 demonstrate the position in which a new image containing the grid class object was extracted. The green, red and blue rectangles represent the cuts for the C.T., panel, and base classes, respectively. Fig. 3 shows two photos taken in the field, the same procedure was performed for the other 30 images.

It is possible to see in Table 1 that the base class has 23 images, and the grid, panel, and C.T. classes have 29, 32, and 32 images, respectively. This was because in some photos captured in the field it was not possible to generate the cuts

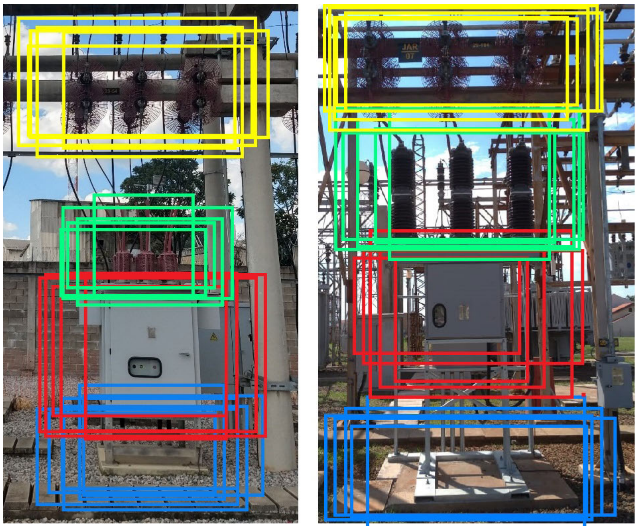

FIGURE 3. Selection of objects for CNN training.

FIGURE 4. Example images contained in the dataset.

for all four classes in the referenced image. For the panel and C.T. classes, in which both objects are centered in the bay, they are present in the 32 photos captured.

Fig. 4 illustrates examples of images with the application of data augmentation that were used in the experiments of this study, aiming to find the best CNN architecture to be part of the proposed system.

TABLE 2. Number of Images Contained in Each of the Five Data Sets.

	Training	Validation	Test
base	680	120	120
grid	800	160	200
panel	800	200	280
c.t.	800	200	280
Total	3080	680	880

There are four classes of objects covered: base, panel, current transformer (C.T.), and power transformer (P.T.). The class ''p.t.'' was called ''grid'' in this work because the object that can be seen in the image is a grid, and not the ''p.t.'' in question. In practice, the grid installed in the bay serves to protect the P.T. from birds that may damage the system. Fig. 4 presents two examples of objects for each of the four classes: base, panel, C.T. and P.T. (grid), in that order. Finally, Table 2 shows the number of images per class for each of the five training, validation, and test sets.

B. IMPLEMENTATION OF CNN ARCHITECTURES

After the stage of preparing the five data sets to be used in the experiments, the implementation of the chosen CNN architectures was carried out. The environment used for coding the methods was Google Colab, which is a cloud service for the development of machine learning algorithms in the Python [37] programming language. Google Colab has some limitations in its free version, making it impossible to run a larger number of experiments and use a more robust set of images. Despite this, its choice is justified by the fact that it supports GPU.

For the implementation of the CNN architectures, ImageAI was used, which is a library for the Python programming language that was developed to enable developers, researchers, and students to create applications and systems with independent resources for deep learning and computer vision [38]. It supports a list of machine learning algorithms for classifying and detecting objects in images, including those used in this work: DenseNet, Inception, ResNet, and SqueezeNet.

The justification for choosing these CNN architectures was based on the work of [39], which discusses the impacts of deep learning and its ability to solve tasks in the area of computer vision. The authors state that a large part of this success in the image classification task is due to the use of convolutional architectures, in which CNNs showed a high classification performance in data sets such as ImageNet, COCO among others. In the same vein, in [40], an analysis of CNN architectures that make up the state of the art was carried out. There are numerous CNN architectures available, and it is not feasible to test all of them for the problem addressed. That said, we tried to select those that obtained satisfactory results in similar tasks, to compare them specifically for our problem. These networks achieved good results in a number of studies that aimed to carry out a comparative study among CNN architectures [41]–[47]

With the available materials, the CNN architectures were implemented respecting their topology as described in the literature, regarding the disposition of neurons in the convolution layers and their respective hyperparameter values,

TABLE 3. Architecture DenseNet-121.

Layer	Output Size	Filter size / stride
conv	112 x 112	7 x 7 / 2
pool	56 x 56	3 x 3 / 2
dense block	56 x 56	1 x 1 (x 6) 3 x 3 (x 6)
transition	56 x 56	1 x 1
	28 x 28	2 x 2 average pool / 2
dense block	28 x 28	1 x 1 (x 12) 3 x 3 (x 12)
transition	28 x 28	1 x 1
	14 x 14	2 x 2 average pool / 2
dense block	14 x 14	1 x 1 (x 24) 3 x 3 (x 24)
transition	14 x 14	1 x 1
	7 x 7	2 x 2 average pool / 2
dense block	7 x 7	1 x 1 (x 16) 3 x 3 (x 16)
	1 x 1	7 x 7 global average pool
		1000-d fully-connected, softmax

TABLE 4. Architecture Inception-v3.

Layer	Input size	Filter size / stride
conv	223 x 223 x 3	3 x 3 / 2
conv	111 x 111 x 32	3 x 3 / 1
conv padded	109 x 109 x 32	3 x 3 / 1
pool	109 x 109 x 32	3 x 3 / 2
conv	54 x 54 x 64	3 x 3 / 1
conv	52 x 52 x 80	3 x 3 / 2
conv	25 x 25 x 192	3 x 3 / 1
3 x Inception	25 x 25 x 192	Inception modules where each 5×5 convolution is replaced by two 3×3 convolution.
5 x Inception	12 x 12 x 1280	Inception modules after the factorization of the $n \times n$ convolutions ($n = 7$ for the 17×17 grid).
2 x Inception	12 x 12 x 1280	Inception modules with expanded the filter bank outputs.
pool	5 x 5 x 2048	8 x 8
linear	1 x 1 x 2048	logits
softmax	1 x 1 x 1000	classifier

obeying the disposition of the network elements, in addition to the adjustment in the feature extraction step.

For all implemented architectures, the image size for the input layer was defined as 224×224. As presented in the work by [23], the DenseNet-121 architecture was chosen for the dense convolutional neural network, as shown in Table 3.

Unlike the other architectures that in their original format already expect a 224×224 image in the input layer, Inception-v3 has a 300×300 size in the input layer, therefore, to maintain the same input proportions as other architectures in this work, Inception-v3 was adapted, according to the work of [24]. Table 4 presents the architecture of the Inception-v3 network used.

For the convolutional neural network that uses a residual learning structure, the ResNet-50 architecture was defined, as shown in Table 5.

Finally, the last CNN implemented was SqueezeNet, and its architecture can be seen in Table 6.

C. EXPERIMENTS

The architectures of CNNs were implemented and for a fair comparison among them, some hyper-parameters were fixed

TABLE 5. Architecture ResNet-50.

Layer	Output size	Filter size / stride
conv	112 x 112	7 x 7, 64 / 2
conv	56 x 56	3 x 3 max pool / 2
		1 x 1, 64 (x 3) 3 x 3, 64 (x 3) 1 x 1, 256 (x 3)
conv	28 x 28	1 x 1, 128 (x 4) 3 x 3, 128 (x 4) 1 x 1, 512 (x 4)
conv	14 x 14	1 x 1, 256 (x 6) 3 x 3, 256 (x 6) 1 x 1, 1024 (x 6)
conv	7 x 7	1 x 1, 512 (x 3) 3 x 3, 512 (x 3) 1 x 1, 2048 (x 3)
	1 x 1	average pool 1000-d fully-connected softmax

TABLE 6. Architecture SqueezeNet.

Layer	Output size	Filter size / stride
conv	111 x 111 x 96	7 x 7 / 2 (x 96)
maxpool	55 x 55 x 96	3 x 3 / 2
fire	55 x 55 x 128	
fire	55 x 55 x 128	
fire	55 x 55 x 128	
maxpool	27 x 27 x 256	3 x 3 / 2
fire	27 x 27 x 256	
fire	27 x 27 x 384	
fire	27 x 27 x 384	
fire	27 x 27 x 512	
maxpool	13 x 13 x 512	3 x 3 / 2
fire	13 x 13 x 512	
conv	13 x 13 x 1000	1 x 1 / 1 (x 1000)
avgpool	1 x 1 x 1000	13 x 13 / 1

during training in each model for each experiment. In all cases, transfer learning from pre-trained weights in the ImageNet data set was used, the number of epochs was 100 and the batch size was 32.

Transfer learning can be defined as the practice of using the knowledge acquired when solving a certain problem, applying it to a different problem. In the experiments carried out, transfer learning was adopted through a pre-trained model in the ImageNet data set, different for each CNN architecture, since each of them has its topology. The advantage of using this type of learning is that the weights of each neuron do not start randomly, reducing the training time. Therefore, each of the architectures has been completely retrained in its entirety, without the freezing of any layer that makes up the network.

The number of epochs is a hyper-parameter that defines the number of times that the learning algorithm will work throughout the training data set. The batch size is a hyper-parameter that defines the number of samples (images) to be analyzed before updating the model's internal parameters.

At the end of each training period, the accuracy obtained is evaluated in the validation set, and then the first epoch model is saved. If the accuracy value is higher than that found in the previous epoch, the model is again saved containing the new parameters obtained in the training of that

FIGURE 5. Free span, having built only the base.

particular epoch. This process is repeated until the previously established number of epochs is completed.

Finally, the test is performed with the latest model, except for images never seen during training, which had previously been separated in the random division between the training, validation, and test sets. The importance of dividing the data into three different sets is the fact that, when using this methodology, it is possible to evaluate with greater precision the capacity that the model has correctly classifying new images from the real world.

In summary, in the first experiment with the first set of data that was randomly divided, each of the four models was trained, validated, and tested. Then, the same process was repeated using the other four sets of images, each of which has a different division for the training, validation, and test sets. With that, it becomes possible to evaluate the generalization for each of the CNN architectures.

D. APPLICATION

Having carried out the experiments with the chosen CNN architectures, it was defined that one that had an attractive value considering the assertiveness and the classification time in the tests performed so that it would be part of the proposed final application.

Whenever the construction of a new bay in a substation begins, support is positioned at a fixed distance of six meters with a networked camera in front of the construction site, as shown in Fig. 5.

Once a day and always at the end of the work shift of the employees responsible for building the bay, an image is captured by the camera attached to the support, which is later sent to the server containing the developed application. With this, the image obtained is processed and analyzed by the chosen CNN model, finding the progress of the work.

To assess the applicability of the proposed system, an experiment was carried out with an image captured by the camera attached to the support at the location where the elements for a new bay will be installed. Fig. 5 exemplifies this image, being a free span, having been built until the moment of capturing the image, only the base.

Before the image is submitted to the CNN architecture and classified, it is pre-processed. Through the OpenCV [48] Computer Vision library, fixed candidate regions are proposed that possibly contain the objects to be classified. As all sizes of each object are known, ten candidate regions are proposed for each of them, varying the size to locate the object in the image, even if there is variation between models of the same object. Also, considering the robustness of the application, if the employee installs the support with the camera not exactly six meters away from the free span, the proposals for regions can also encompass the object.

Having defined the ten proposals for regions for each object in the image obtained, they are cropped, resulting in a total of 40 images. It is known exactly which object each of the 40 proposed regions should belong to. With this, the ten proposals of regions that should contain the ''base'' object are first evaluated. If the majority of the ten images are classified as ''base'' with a probability of confidence of the model greater than 80%, the object was constructed properly.

Under the condition of classification of the ''base'' object, the classification of the objects ''panel'', ''c.t.'' and ''grid'' is made, respectively. If an object in this respective order is not classified as expected, the other proposals of regions cut out from the original image are not evaluated, since, in the construction of the bay, it is not possible to install a P.T., without having previously installed the panel, for example.

Following such procedures, a daily log is generated with all the information obtained, which can be accessed at any time by the managers who are responsible for monitoring the constructions, about the construction of bays in electrical substations, which by chance have the proposed system installed. Also, it is possible to configure e-mails for the automatic receipt of these logs, at the determined frequency.

V. RESULTS AND DISCUSSION

A. CNN ARCHITECTURE CHOICE

The results presented are based on two premises: (1) the model saved at the time when it converged and obtained the highest accuracy value before the validation set; (2) the number of correct answers for each of the classes in the images contained in the test set. For each of the five different data sets, care was taken that the test set had 120 images of the base class, 200, 280, and 280 for the grid, panel, and c.t. classes, respectively.

In general, it was noticed that the training time of the CNN SqueezeNet architecture was much shorter than the others, which is justified by having a leaner architecture, aiming at

TABLE 7. Average Results Obtained in the Experiments Carried Out.

	DenseNet	Inception	ResNet	SqueezeNet
Epochs	5	6	15	12
base	100%	99.66%	100%	99.66%
grid	97.30%	100%	100%	100%
panel	100%	98.85%	98%	99.07%
c.t.	88.57%	100%	99.85%	100%
Average	96.46%	99.63%	99.46%	99.68%

applications that have a lower hardware capacity, such as embedded devices, for example.

Because of transfer learning, since it was decided to use pre-trained models in the ImageNet data set, the number of iterations (epochs) to converge the models (regardless of architectures) was low, especially for DensetNet and Inception, which are more robust CNN architectures. If the network weights had been initialized randomly, it is believed that the number of iterations for converging the models would have been much higher.

In the number of hits, regardless of the random division of the training, validation, and test sets, both CNN architectures behaved very well, with an accuracy of almost 100% in the test images. Also, it was found that for the errors obtained, the probability of the class with the highest value was, almost always, approximately 40%, that is, the model obtained two probabilities (one for the correct class and another for the class that it missed) very close and low, which caused the classification error.

Table 7 shows the average results obtained in the five experiments carried out on each set of different images, for each of the CNN architectures addressed.

It is possible to observe in Table 7 that the average accuracy of all objects was almost 100%, regardless of the CNN architecture used. The DenseNet classifier had worse performnace in the "c.t." class, in relation to other competing architectures.

When specifically analyzing the predicted probabilities of the images that were incorrectly classified by the CNN SqueezetNet architecture in the experiments carried out, it was seen that not all images had the highest probability predicted in just a certain class, that is, the "panel" object was not only confused with the "base" object in all errors, for example. The erroneously classified images had, on average, a value of approximately 33% for the classified object (which may be base, grid, or c.t.) and an approximate value for the panel class.

To statistically compare the results obtained, the Shapiro-Wilk test was applied to the four distributions and it was verified that the data follow a normal distribution. Given this premise, the Analysis of Variance test (ANOVA) was used and, given a significance level of 0.05, there was the acceptance of the null hypothesis and rejection of the alternative hypothesis, suggesting that there is no difference between the performance in the experiments with each architecture.

As there was no significant statistical difference between the results obtained, we chose the CNN SqueezetNet architecture, since its training time was much shorter than the others, which makes sense knowing that its architecture is simpler, and yet it achieved high accuracy. In addition, because of its simplicity, the classification time for new images is also much shorter.

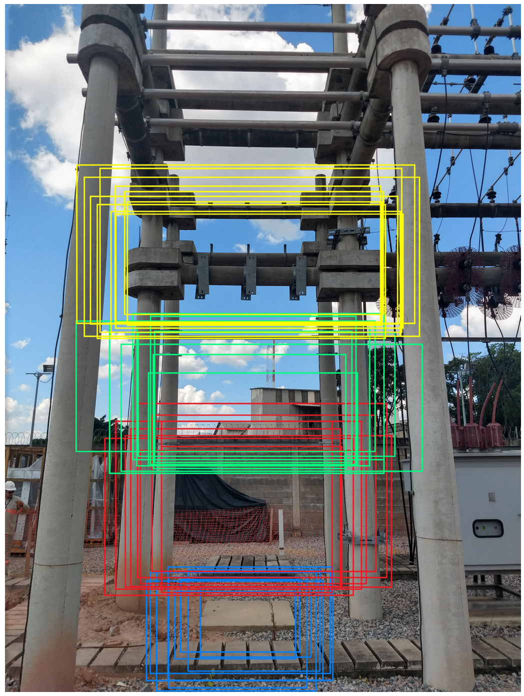

FIGURE 6. Proposals from candidate regions for each object.

B. PROPOSED APPLICATION

Fig. 6 presents an image of a free span captured by the camera attached to the support, which was installed at six meters and in front of the place where the bay will be built. The colored rectangles represent the proposals of candidate regions for each of the four objects covered in the proposed application, in the image pre-processing stage.

It is expected that the blue rectangles will be able to capture the base object, while the red, green, and yellow rectangles, the panel, C.T. and grid objects, respectively. Regardless of the variation of the models of a given object and, in case the fixation of the support with the camera is installed in the non-corresponding location, the variability of the sizes and positions of the proposals of regions manage to circumvent these possible problems.

Table 8 (left) shows the average result of the probabilities obtained by classifying the 10 cutouts of the blue color illustrated in Fig. 6. As expected, since there is a base built on the site, the model was able to classify correctly with a high precision value.

Knowing that the blue cutouts illustrated in Fig. 6 can only be of the base class, it is expected that the classifier will be able to find that same object. As the conditions of most of the ten images were classified as "base" and with a probability of confidence of the model greater than 80%, the next object must be evaluated. Table 8 (right) presents the average result

TABLE 8. Results Obtained in the Experiment With a Free Span for the Base and Panel Class.

Expected class: **base**		Expected class: **panel**	
Class	**Probability**	**Class**	**Probability**
base	**94.25%**	base	48.27%
panel	0.84%	**panel**	**22.74%**
grid	1.18%	grid	10.91%
c.t.	3.73%	c.t.	18.08%

of the probabilities obtained by the classifications of the 10 cutouts of the red color illustrated in Fig. 6.

It was expected that the images obtained through the red cutouts would be classified as belonging to the object "panel", however, as the gap is free and a panel has not yet been installed, the classifier was unable to find it, returning low probabilities for the classified images. As a result, if the conditions of most of the ten images are not classified as "panel" and with a model's probability of greater than 80%, the system will not try to find the elements of the bay that are still to be built.

In this way, it becomes possible to provide the managers responsible for managing the construction of bays in electric power substations, an automatic method of monitoring, so that such managers no longer need to move to each construction in each substation, since that it is quite common for the same company to have facilities in different cities or even different states.

VI. CONCLUSION

In this work, a system was proposed that allows the automatic monitoring of the management of the constructions in electric power substation bays. A comparative study was carried out between four CNN architectures: DenseNet, Inception, ResNet, and SqueezeNet applied to a classification problem of four elements that make up bays in substations. From 32 field photos of different bays captured in substations, thousands of images of each component class were obtained. Subsequently, data augmentation operations were applied in order to obtain a larger set for training. Five different sets of images were randomly generated, each divided into training, validation, and test sets.

Regardless of the CNN architecture employed, it is concluded that the models obtained during the experiments have a high assertiveness for the problem in which they were applied, approaching 100%. It was observed that there is a difference in relation to the CNN SqueezeNet architecture, which has a shorter training time and occupies much less disk space compared to the others and yet managed to obtain a similar assertiveness in the test images.

Consequently, other CNN architectures can be compared for the same problem, and, as future work, other more complex classification problems can also be applied and used for comparison. The current comparative study was a fundamental piece for the implementation of an object classification method that uses CNNs to extract features from the input images and then classifies each object in the image since it has already been pre-processed by using proposals from candidate regions.

One of the main problems encountered in the development of the work was the collection of images. Due to the dangerous nature of electrical substations, any visitor must be accompanied by a qualified professional, also, all due care must be taken about the use of personal protective equipment. This situation was a direct obstacle so that there was no way to obtain a larger number of images from different bays.

Another obstacle that has been identified is the limitation of computational resources to train a convolutional neural network. The training time is quite considerable when powerful hardware is not available. Therefore, it was not possible to carry out a greater number of experiments.

In addition to covering the comparative study with other CNN architectures for the classification task, one can approach the problem as an object detection task, exploring methods such as YOLO, Faster R-CNN, SSD among others. With this, it becomes possible to further increase the independence regarding the importance of installing the support with the camera in the correct location in front of the free span, in addition to being plausible to obtain exact information on the location of the object in the image.

In addition, as future work, a web application will be developed with attractive layouts for the manager, so that they can obtain more detailed information than that contained in the logs generated by the current proposed application. In a later system, it is possible, for example, to allow the manager to be able to assist in real-time the execution of a specific work at any time, in addition to the implementation of data analysis methods for the generation of intelligent reports, proposing alternatives for managers, through the data obtained.

REFERENCES

[1] G. P. J. Verbong and F. W. Geels, "Exploring sustainability transitions in the electricity sector with socio-technical pathways," *Technol. Forecasting Social Change*, vol. 77, no. 8, pp. 1214–1221, Oct. 2010.

[2] W. Wang and K. Siau, "Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda," *J. Database Manage.*, vol. 30, no. 1, pp. 61–79, Jan. 2019.

[3] P. de Oliveira e Lucas, M. A. Alves, P. C. de Lima e Silva, and F. G. Guimarães, "Reference evapotranspiration time series forecasting with ensemble of convolutional neural networks," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105700.

[4] G. Currie, K. E. Hawk, E. Rohren, A. Vial, and R. Klein, "Machine learning and deep learning in medical imaging: Intelligent imaging," *J. Med. Imag. Radiat. Sci.*, vol. 50, no. 4, pp. 477–487, Dec. 2019.

[5] A. Braun and A. Borrmann, "Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning," *Autom. Construct.*, vol. 106, Oct. 2019, Art. no. 102879.

[6] G. T. Heydt, C. C. Liu, A. G. Phadke, and V. Vittal, "Solution for the crisis in electric power supply," *IEEE Comput. Appl. Power*, vol. 14, no. 3, pp. 22–30, Jul. 2001.

[7] G. Bonnard, "The problems posed by electrical power supply to industrial installations," *IEE Proc. B, Electr. Power Appl.*, vol. 132, no. 6, pp. 335–343, Nov. 1985.

[8] J. Furman, J. P. Holdren, C. Muñoz, M. Smith, and J. Zients, "Artificial intelligence, automation, and the economy," Executive Office President, Washington, DC, USA, Tech. Rep., Dec. 2016.

[9] D. Autor and A. Salomons, "Is automation labor-displacing? Productivity growth, employment, and the labor share," Nat. Bur. Econ. Res., Cambridge, MA, USA, Working Paper 24871, Jul. 2018. [Online]. Available: http://www.nber.org/papers/w24871, doi: 10.3386/w24871.

[10] O. Gomes and S. Pereira, "On the economic consequences of automation and robotics," *J. Econ. Administ. Sci.*, vol. 36, no. 2, pp. 134–153, Jan. 2019.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[14] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3106–3121, Dec. 2019.

[15] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, Jun. 2014.

[16] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Feb. 2018.

[17] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[18] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. 1. Reading, MA, USA: Addison-Wesley, 1992.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[20] T. N. Sainath, A.-R. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8614–8618.

[21] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 994–1000.

[22] N. Kriegeskorte, "Deep neural networks: A new framework for modeling biological vision and brain information processing," *Annu. Rev. Vis. Sci.*, vol. 1, no. 1, pp. 417–446, Nov. 2015.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[26] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360

[27] Y. Lin, J. Sun, C. Li, Y. Ma, Y. Geng, and Y. Chen, "Deep learning for intelligent substation device infrared fault image analysis," in *Proc. MATEC Web Conf.*, vol. 55. Les Ulis, France: EDP Sciences, 2016, p. 03007.

[28] C.-Z. Fu, W.-R. Si, H. Huang, L. Chen, Q.-J. Gao, C.-B. Shi, and C. Wang, "Research on a detection and recognition algorithm for high-voltage switch cabinet based on deep learning with an improved YOLOv2 network," in *Proc. 11th Int. Conf. Intell. Comput. Technol. Autom. (ICICTA)*, Sep. 2018, pp. 346–350.

[29] J. Wang, "Deep learning based state recognition of substation switches," in *Proc. AIP Conf.*, 2018, vol. 1971, no. 1, Art. no. 040041.

[30] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Autom. Construct.*, vol. 85, pp. 1–9, Jan. 2018.

[31] H. Kim, H. Kim, Y. W. Hong, and H. Byun, "Detecting construction equipment using a region-based fully convolutional network and transfer learning," *J. Comput. Civil Eng.*, vol. 32, no. 2, Mar. 2018, Art. no. 04017082.

[32] W. Fang, L. Ding, B. Zhong, P. E. D. Love, and H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Adv. Eng. Informat.*, vol. 37, pp. 139–149, Aug. 2018.

[33] S. Arabi, A. Haghighat, and A. Sharma, "A deep learning based solution for construction equipment detection: From development to deployment," 2019, *arXiv:1904.09021*. [Online]. Available: http://arxiv.org/abs/1904.09021

[34] Y. Liu, J. Liu, and Y. Ke, "A detection and recognition system of pointer meters in substations based on computer vision," *Measurement*, vol. 152, Feb. 2020, Art. no. 107333.

[35] I. Ullah, R. U. Khan, F. Yang, and L. Wuttisittikulkij, "Deep learning image-based defect detection in high voltage electrical equipment," *Energies*, vol. 13, no. 2, p. 392, Jan. 2020.

[36] A. Jadia and M. Chawla, "Image classification and detection of insulators using bag of visual words and speeded up robust features," *Int. J. Innov. Sci. Modern Eng.*, vol. 6, pp. 7–13, Sep. 2020. [Online]. Available: https://www.ijisme.org/wp-content/uploads/papers/v6i10/J12600961020.pdf

[37] E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Berlin, Germany: Springer, 2019.

[38] O. J. Moses, "ImageAI, an open source python library built to empower developers to build applications and systems with self-contained Computer Vision capabilities," Tech. Rep., Mar. 2018. [Online]. Available: https://github.com/OlafenwaMoses/ImageAI

[39] L. Jiao and J. Zhao, "A survey on the new generation of deep learning in image processing," *IEEE Access*, vol. 7, pp. 172231–172263, 2019.

[40] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, pp. 64270–64277, 2018.

[41] N. Ghassemi, H. Mahami, M. Tayarani Darbandi, A. Shoeibi, S. Hussain, F. Nasirzadeh, R. Alizadehsani, D. Nahavandi, A. Khosravi, and S. Nahavandi, "Material recognition for automated progress monitoring using deep learning methods," 2020, *arXiv:2006.16344*. [Online]. Available: http://arxiv.org/abs/2006.16344

[42] D. Gil, G. Lee, and K. Jeon, "Classification of images from construction sites using a deep-learning algorithm," in *Proc. 35th Int. Symp. Autom. Robot. Construct. (ISARC)*, Jul. 2018, pp. 1–6.

[43] B. Espejo-Garcia, N. Mylonas, L. Athanasakos, and S. Fountas, "Improving weeds identification with a repository of agricultural pre-trained deep neural networks," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105593.

[44] A. R. Saikia, K. Bora, L. B. Mahanta, and A. K. Das, "Comparative assessment of CNN architectures for classification of breast FNAC images," *Tissue Cell*, vol. 57, pp. 8–14, Apr. 2019.

[45] M. H. Saleem, J. Potgieter, and K. M. Arif, "Plant disease classification: A comparative evaluation of convolutional neural networks and deep learning optimizers," *Plants*, vol. 9, no. 10, p. 1319, Oct. 2020.

[46] Ananda, C. Karabag, A. Ter-Sarkisov, E. Alonso, and C. C. Reyes-Aldasoro, "Radiography classification: A comparison between eleven convolutional neural networks," in *Proc. 4th Int. Conf. Multimedia Comput., Netw. Appl. (MCNA)*, Oct. 2020, pp. 119–125.

[47] K. Meshkini, J. Platos, and H. Ghassemain, "An analysis of convolutional neural network for fashion images classification (Fashion-MNIST)," in *Proc. Int. Conf. Intell. Inf. Technol. Ind.* Cham, Switzerland: Springer 2019, pp. 85–95.

[48] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision With the OpenCV Library*. Newton, MA, USA: O'Reilly Media, 2008.

BRUNO ALBERTO SOARES OLIVEIRA received the bachelor's degree in computer engineering from the Federal Institute of Education, Science, and Technology of Minas Gerais - Campus Bambuí (IFMG), in 2018, and the master's degree in electrical engineering the Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, in 2020, in the line of research in computational intelligence, where he is currently pursuing the Ph.D. degree.

He was an Associate Member of the Center for Research in Law, Technology, and Innovation (Centro DTIBR). He became a Technician in internet computing from the Educational Institution of Arcoense (INPA), in 2012. Since February 2020, he has been a full Developer of intelligent systems using machine learning and computer vision with SVA Tech - Analytical Video Solutions. He is a member of the Machine Intelligence and Data Science Laboratory (MINDS) for computational intelligence research. His research interests include machine learning algorithms, data science, computer vision, and digital image processing.

ABÍLIO PEREIRA DE FARIA NETO received the bachelor's and master's degrees in computer science from the Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, in 1986 and 1991, respectively.

He was with Fitec, from 2001 to 2014, Lucent Technologies, from 1999 to 2001, Batik, from 1994 to 1999, Audiolab, from 1988 to 1994, and Malc Automação, from 1986 to 1988. Since 2015, he has been a Manager and a Software Developer with SVA Tech - Soluções em Vídeo Analytical. He has more than 30 years of experience in software development, with an emphasis on basic software, embedded software, and communication protocols. More than 20 years of experience in managing teams and software processes. He has authored participated in several international projects with distributed and multicultural teams. His research interest includes computer vision.

ROBERTO MÁRCIO ARRUDA FERNANDINO was with Sony da Amazônia with video cameras, where he was also with the Assembly Line and Technical Maintenance Department, authorized network management nationwide, from 1992 to 1998. He was a Founding Partner of Veotex Sistemas Tecnológicas SA, where he was the Technical Director responsible for P&D Management, from 2003 to 2013. He is currently a Founding Partner of SVA Tech - CEO of the company, responsible for the management, technical area, and training of models using high-performance neural netconstructions. Has experience in computer science, with emphasis on methodology and computer techniques. More than 30 years of experience in the field of imaging systems.

ROGÉRIO FERNANDES CARVALHO received the bachelor's degree in computer science and the master's degree in electrical engineering from the Pontifical Catholic University of Minas Gerais, Belo Horizonte, Brazil, in 1998 and 2011, respectively.

He was a Software Development Specialist for more than 10 years at Fitec-Inovações Tecnológicas. During this period, he participated in projects developed in partnership with Alcatel-Lucent's Bell Labs, also operating abroad (Phoenix-AZ and Marlboro-MA, USA). Since 2015, he has been the Director and a Senior Software Developer with SVA Tech - Soluções em Vídeo Analytical. He participated in the Research Group in Parallel and Distributed Processing, carrying out research and development activities (P&D) with the application of high-performance processing to solve important problems in large areas of knowledge. His current research interest includes computer vision. He had the opportunity to participate in multicultural teams around the world developing software according to the strict quality standards required by Bell Labs.

AMANDA LOPES FERNANDES received the title of technologist in Foreign Trade from Universidade Paulista (UNIP) and a Specialist in supply chain and logistics management from the Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil, in 2013 and 2017, respectively.

Since August 2019, she has been a full Analyst of innovation projects with CPFL Energia. She manages and participates in research and development projects through the P&D program ANEEL whose proponent is CPFL Paulista. She has seven years of experience in logistics and project management. She worked in computer vision projects, digital image processing, software development, generation of new methodologies, and creation of new products.

FREDERICO GADELHA GUIMARÃES (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in electrical engineering from the Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, in 2003, 2004, and 2008, respectively.

In 2010, he joined the Department of Electrical Engineering, UFMG, where he became an Associate Professor in 2018. Since 2014, he has been responsible for the Machine Intelligence and Data Science Laboratory (MINDS) for computational intelligence research. He also had a Postdoctoral Fellowship with the Laboratoire Images, Signaux et Systèmes Intelligents (LiSSi), Université Paris-Est Créteil (UPEC), Paris, France, from 2017 to 2018. He has published more than 200 articles in journals, congresses, and chapters of national and international books. He has experience in electrical engineering and computer engineering, with an emphasis on optimization, computational intelligence, machine learning, time series forecasting, genetic algorithms, and evolutionary computation.

Dr. Guimarães is a member of the IEEE Computational Intelligence Society (CIS) and the IEEE Systems, Man, and Cybernetics Society (SMCS). He is also an Associate Editor of the journals IEEE ACCESS and *Neurocomputing* (Elsevier).