COBEM 2021
26th International Congress of Mechanical Engineering

ABCM
Associação Brasileira de Engenharia e Ciências Mecânicas
November 22-26, 2021, Virtual Congress, Brazil

COB-2021-0981
DEEP CONVOLUTIONAL NEURAL NETWORKS FOR IMAGE CLASSIFICATION: A CASE STUDY IN AN ELECTRIC UTILITY WAREHOUSE

Paulo Henrique Martinez Piratelo
Electrical Engineering Graduate Program (PPGEE), Federal University of Parana (UFPR), Curitiba, Brazil
Mechanical Systems, Lactec Institute, Curitiba, Brazil
paulo.piratelo@lactec.org.br

Rodrigo Negri de Azeredo
Mechanical Systems, Lactec Institute, Curitiba, Brazil
rodrigo.azeredo@lactec.org.br

Eduardo Massashi Yamao
eduardo.yamao@lactec.org.br

Gabriel Maidl
gabriel.maidl@lactec.org.br

Rafael Martini Silva
rafael.martini@lactec.org.br

Laercio Pereira de Jesus
Dept. of Logistics and Supplies, Copel Distribuição S.A, Curitiba, Brazil
ldejesus@copel.com

Renato de Arruda Penteado Neto
Mechanical Systems, Lactec Institute, Curitiba, Brazil
renato@lactec.org.br

Leandro dos Santos Coelho
Industrial and Systems Eng. Graduate Program (PPGEPS), Pontifical Catholic University of Parana (PUCPR), Curitiba, Brazil
Electrical Engineering Graduate Program (PPGEE), Federal University of Parana (UFPR), Curitiba, Brazil
leandro.coelho@pucpr.br

Gideon Villar Leandro
Electrical Engineering Graduate Program (PPGEE), Federal University of Parana (UFPR), Curitiba, Brazil
gede@ufpr.br

Abstract. *Warehouse management has proven to be fundamental for improvements in the productivity and organization of companies, bringing several benefits in controlling the flow of products and operations. As convolutional neural networks (CNNs) are great tools of deep learning to operate classification tasks in computer vision, six models of architectures were tested in a classification assignment through a red-green-blue (RGB) image dataset to classify equipment, allocated in the warehouse of an electric utility. Thus, the present work aims to compare the performance of these models in the asset identification process. The dataset consisted of 565 images was built in local, in an uncontrolled environment, representing real challenges that occur in many warehouses. SqueezeNet obtained the best results, reaching an accuracy of 97.5% and a F1 score of 97.4% in the test set. This comparison can be used in order to guide future works on automated and intelligent inventory management solutions on the electric maintenance field.*

Keywords: *convolutional neural networks, deep learning, computer vision, warehouse management, electric utility.*

1. INTRODUCTION

Researches conducted on literature shows that in the last decades there were several studies on the theory of inventory management improvements and new technologies that could be applied in warehouse management systems (Yang *et al.*, 2021). The warehouse management performs a vital role in industrial productivity, operational status, level of readiness and customer service (Al-Momani *et al.*, 2020; Tejesh and Neeraja, 2018), controlling the flow of products, time and operational cost. Moreover, in the last decade, Artificial Intelligence (AI) played an important function in the supply chain management field, with customer demand predictions, order fulfillment, and picking. However, it is reported a lack of study on the warehouse receiving stage (Yang *et al.*, 2021).

A case study with 500 enterprises shows that AI helped in manufacturing, automating planning, and inventory, eliminating inaccurate and time-consuming processes and improving real-time visibility of assets (Wamba-Taguimdje *et al.*, 2020). A research points that the profits from AI have the potential to increase profitability by 38%, boosting US$14 trillion across 16 industries by 2035 (Purdy, M. and Daugherty, P., 2017). Furthermore, a global contribution of US$6 trillion from increasing productivity and US$9 trillion from consumption-side effects could be reached by 2030 (PwC, 2017). However, a digital transformation is not enough to reach intelligent solutions. Data needs to be transformed into knowledge in order to achieve the benefits. Advanced analytics and intelligent algorithms combined with human intelligence are key to a positive impact for the organization (Lichtenthaler, 2020).

Thus, this transformation of data into information for guided decision-making can take place in several ways. In the case of inventory management, images of item layout can be used to recognize objects and control the organization of the warehouse. For this activity, an image classifier is used, extracting features from the images through different mathematical techniques and classifying the objects. There are several image classification techniques such as Convolutional Neural Network (CNN) and their different architectures variations, besides those other methods such as Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Random Forest Algorithm and Particle Swarm Optimization (PSO), as shown in Sanghvi *et al.* (2020); Preet Kour and Arora (2020); Zhang *et al.* (2020).

CNNs (LeCun *et al.*, 1995) have been used in computer vision applications, such as classification of images, video processing, natural language processing, and segmentation. They are a specific type of neural network and have a powerful capacity for learning. CNNs use multiples feature detectors that automatically learn data representation (Khan *et al.*, 2020). Residual Neural Networks (Resnets) have been used in warehouses for tasks like electronic parts classification (Patel and Chowdhury, 2020) and improving localization with the help of a camera, sensors, and liDAR (Relyea *et al.*, 2020). A Deep Convolutional Neural Network (DNN) architecture called AlexNet is a base foundation for an application in a logistics sorting warehouse application (Chen and Dong, 2021). In the electrical maintenance field, the CNN from Visual Geometry Group of Oxford (VGG) is combined among different methods in the detection of electric towers in complex environments (Tian *et al.*, 2020). An architecture called SqueezeNet was modified and used for the task of product recognition in e-commerce recommendation scenes, improving accuracy in image classification (Fan *et al.*, 2020). A proposed method using reusable feature maps was applied in a Densely Connected Convolutional Network (DenseNet) in order to attack the task of remote sensing scene classification, improving state-of-the-art performance (Zhang *et al.*, 2019).

In consequence, a Brazilian electrical company wants to automate the process of inventory control and material flow. The company is seeking AI technologies applied with computer vision and automation in order to improve warehouse management. This present work focuses on the comparison of some CNNs applied in a real environment of an electric utility warehouse. The applied models are Resnet, AlexNet, VGG, SqueezeNet, DenseNet, and Inception. The CNNs have the role to classify electric parts allocated in the warehouse through a red-green-blue (RGB) image dataset. The dataset was built-in local, in an uncontrolled environment, and represents a challenge that occurs in real applications.

The remainder of this article is structured as follows. Section 2 approaches the problem description, explaining the procedures to capture the images and dataset composition. Section 3 presents the related works developed in the area. Then, Section 4 deals with the deep learning models used in the research to meet the requirements of the project. In Section 5 the results are presented and discussed. Finally, Section 6 brings the conclusion and future works.

2. PROBLEM DESCRIPTION

The electrical utility warehouse where the dataset was built is an 11,000 square meters building, with more than 3,000 types of objects related to electrical power systems maintenance, distributed in shelves across the entire facility. The company is facing management issues that need to be addressed such as miscounting, high time-consuming processes, and costs of inventory control by manual checking and material flow. It is intended to mitigate these issues by applying new technology and improve the process in inventory management. One of the ways to attack these problems is an automated inventory check, based on artificial intelligence, computer vision, and automation. It is proposed to build a tool that classifies the products inside the warehouse using techniques of deep learning. The tool recognizes the objects displayed on the shelves through RGB images captured by an automated guided vehicle (AGV).

The warehouse presents a challenge for the computer vision and machine learning application, due to the fact that it is an uncontrolled environment. There are no sufficient rules applied to the displacement of materials. Consequently, the tool needs to be robust enough to overcome this issue. Moreover, the lighting is also a concern since there are some locations with the absence of light. Thus, a dataset was built in this environment, which represents a real challenge that occurs in many similar places of inventory storage. The dataset consists of RGB images of two different classes of objects: utility pole insulators and brace bands. Figure 1 shows some examples of the images that are captured where (a) are the pole insulators and (b) brace bands.

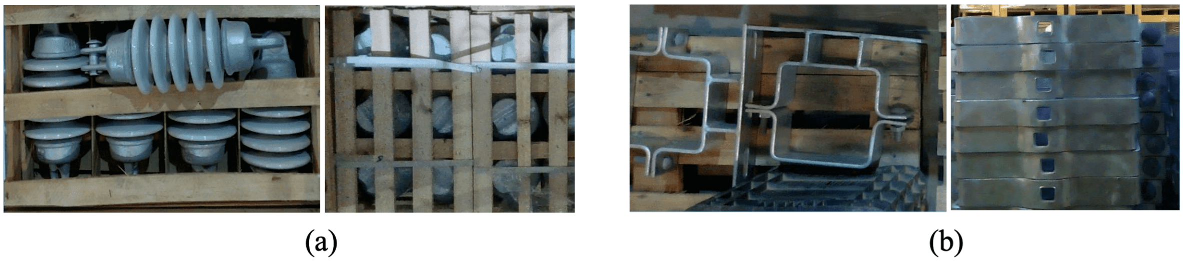

Figure 1. Dataset images examples: (a) pole insulators and (b) brace bands.

3. RELATED WORKS

Considering this scenario and the assets management of companies in the energy power sector, a search was carried out to verify existing works in this field, where deep learning and computer vision are applied to recognize electrical devices. A faster region based convolutional neural network (Faster R-CNN) was used to detect equipment in an electric power room (Zhang *et al.*, 2021). For that experiment, a 5600 image dataset of 100 different classes was built and manually labeled. A random shuffle was applied in the dataset, dividing it into 70% of images for training the net, 10% for tuning it in a validation set, and 20% composing the test set, in order to verify the model. The experiment achieved 91.3% mean average precision (mAP) on the test set. However, the authors pointed out a difficulty in detect dense small objects, which is exactly the case for the application on this electric utility warehouse.

A dataset consisting of 804 pictures was built for an electric power equipment image recognition application based on a deep forest learning model (Yao and Cheng, 2021). There are 5 classes on this balanced dataset, being insulators, transformers, circuit breakers, poles, and iron towers. The images have different sizes. The dataset received an expansion, including images of civil equipment in order to verify the algorithm's ability to recognize the objects. In a study conducted by Lile and Yiqun (2017), it was used a VGG architecture, and the results showed its capacity of detecting anomalies in electrical equipment with images of infrared thermography as input for the net. The thermography dataset is divided into four categories, being bus bar (meter), bus bar (circuit), IP phone, and PC motherboard. The dataset consists of 1140 images for training, 285 images for validation, and 142 images for the test. In a batch, 50% of images in the training set received a horizontal/vertical flip for data augmentation.

A pre-trained Resnet50 was applied in the task of classifying mechanical parts that enter or leave a smart factory facility (Patel and Chowdhury, 2020). The transfer learning technique fine-tuning was used to transfer the knowledge from the pre-trained model to the desired task, adjusting all the layers of the net. The dataset consists of 1427 images of 6 different mechanical parts, divided into 955 samples for training and 472 for testing. For the training set, the images received a data augmentation (shift, rotation, zoom, reflection, contrast, and brightness) to gather more data and get a better result. The fine-tuned model achieved an accuracy of 98.94% on the test set.

Image recognition and processing model based on a combination of CNN with a recurrent neural network (RNN) as an encoder-decoder were proposed to detect electrical equipment and generate English sentences, describing the scene and helping in inspection (Xia *et al.*, 2018). The dataset was created with the help of on-site patrol teams. A second dataset was built, applying data augmentation by flipping, rotating, and scaling the images. A transfer learning technique was used for the task of adjusting the weights of the net. The feature extraction technique tuned only the last layers of the net, using three pre-trained models as reference (VGG-16, VGG-19, and Inception-V3). Inception-V3 showed better performance accuracy. The authors tested the dataset, before and after the data augmentation, and in 93% of the tests the models had a better recognition rate after the data augmentation.

According to these related papers, there are some procedures and techniques that are commonly applied in the recognition task of electrical parts and classification of objects inside warehouses. Transfer learning techniques was identified to overcome problems like high computational cost, time-consuming learning process and overfitting. Fine-tuning and feature extraction were used to adjust previous knowledge from pre-trained models for new applications. It was common to use a random shuffle on the samples for splitting the dataset, or k-fold cross-validation to test different combinations of samples, both techniques for prevention of biased results. Small datasets also received a data augmentation, by adding more samples from different sources or by changing some existing images. This information guides our experiments on testing the proposed architectures.

4. DEEP LEARNING MODELS

In this section, the deep learning models applied to classify the images of objects allocated in the electric power company's warehouse will be presented, describing each of the neural networks, their properties, and particularities. For this application, the networks used were pre-trained models with the ImageNet dataset. The models received a transfer learning technique (Torrey and Shavlik, 2010) called feature extraction, which takes advantage of knowledge from previous training and applies it to a specific task. It works by updating the final layer of the architecture, reshaped from a thousand neurons (ImageNet classes) to only two (insulator and brace band). Then, the biases and weights of the last layer are trained on the new dataset, leveraging its knowledge to classify the new objects. The networks were implemented through the deep learning library PyTorch [1].

4.1 AlexNet

AlexNet was the first CNN that had an enormous impact on image recognition. The proposed architecture won the ILSVRC-2012 competition with a Top-5 error rate of 15.3% on test set (Krizhevsky *et al.*, 2012). The model was trained on subsets of ImageNet that were used in the ILSVRC-2010 and ILSVRC-2012 competitions. Among the many contributions of this paper, the main one is the use of GPU for training the model. The use of GPU delivered the capacity of training deep models in bigger datasets.

Mainly, the architecture of AlexNet consists of five convolutional layers and three fully connected layers, the last one being a soft-max function. Figure 2 shows the AlexNet architecture. The first convolutional layer has 96 feature maps. The second layer is also a convolutional one with a number of 256 feature maps. The third convolutional layer has 384 feature maps and it does not have a max pooling. The output goes directly to the fourth convolutional layer, which is built exactly like the third one. The fifth layer is a convolutional one with 256 feature maps. The next layer is a fully connected layer with 4096 neurons, followed by another fully connected layer of the same size. At last, the third fully connected layer of 1000 neurons has the soft-max function, in contrast with all the previous layers with rectified linear units (ReLU) as the activation function.

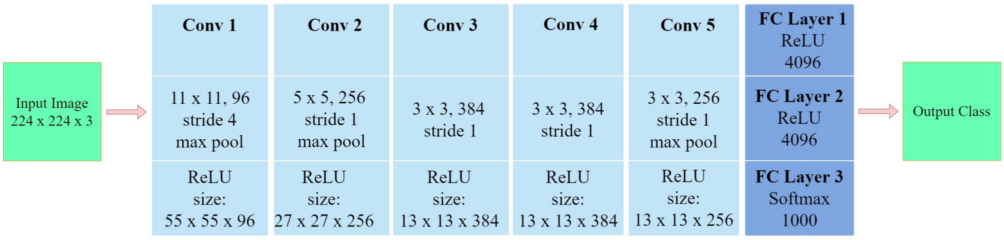

Figure 2. AlexNet architecture.

The architecture has a local response normalization, located after the ReLU function in between convolutional layer one and two and convolutional layer two and three. The authors implemented this function in order to give more generalization to the local normalization scheme. The term $a^i_{x,y}$ is the neuron's activity when given a certain kernel i at the specific position x and y. The variable N is the total number of kernels while n is the number of adjacent kernel maps at the same spatial position. There are three constants as well, being k, α, and β.

The response normalized activity $b^i_{x,y}$ follow the Eq. (1)

$$b^i_{x,y} = a^i_{x,y} \Big/ \left(k + \alpha \sum_{j=max(0,i-n/2)}^{min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^{\beta} \tag{1}$$

The constants k = 2, n = 5, α = 10^{-4} and β = 0.75 are hyperparameters determined in the validation set. The authors also brought to the table the use of data augmentation and dropout. For data augmentation, it was proposed to do image translations, horizontal reflection, and intensity alteration in RGB channels. Dropout was used in all neurons of fully connected layers one and two, multiplying their output by a factor of 0.5.

4.2 VGGNet

Proposed by Simonyan and Zisserman (2014), the VGG architecture is a convolutional neural network model that achieved 92.7% on top-5 test accuracy in ILSVRC. The architecture model used was the VGG-11 with batch normalization, with 11 weight layers in the network (8 convolutional and 3 fully-connected layers), as shown in Fig. 3. The network

[1] https://pytorch.org/

input has a fixed size of 224x224 RBG image. Firstly, pre-processing is performed, where the average RGB value is subtracted from each pixel of the images in the training set. Then, the image crosses several convolutional layers with 3 x 3 dimensional filters.

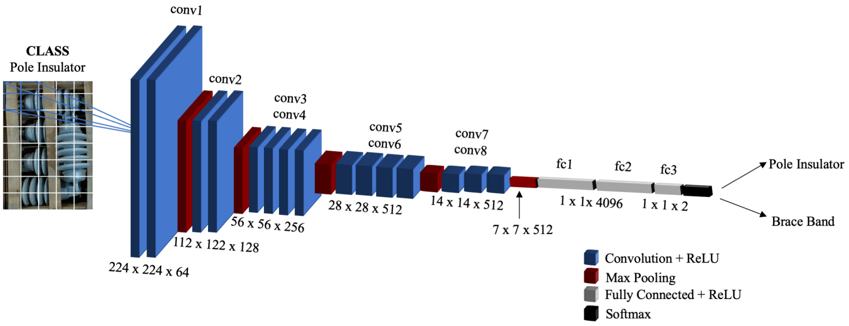

Figure 3. VGG-11 model architecture.

As described by Simonyan and Zisserman (2014), spatial pooling is carried out by five max-pooling layers. The max-pooling is executed in a 2 × 2 pixel window, with stride 2. The width of the convolutional layers (number of channels) starts with 64 in the first layer and grows with a factor of 2 after each max-pooling layer, reaching the end with 512 channels. Next, there are three Fully-Connected (FC) layers, where the first two have 4096 channels and the third has 1000 channels to make the classification of the ILSVRC, i.e. one for each class of the original problem. In this research, since it is a binary problem, the value was adjusted to two classes. This final layer is a soft-max layer.

Consequently, the approximate number of parameters in this network is 133 million. Moreover, the training set receives a data augmentation with the horizontal rotation and RBG shifting randomly of the cropped images. Thus, training from scratch for this network can be very slow and require a considerable computational cost. A transfer learning of the feature extraction type was carried out to reuse the net weights and only the last layer was adjusted for the proposed approach.

4.3 Inception

The Inception architecture (Szegedy *et al.*, 2015) brought the idea of not only deeper but also wider nets. The proposed architecture works with 3 filters of different sizes, a 1 squared filter, a 3 x 3, and finally, a filter with a size of 5 in both dimensions. The second module of Inception deals with a reduction of dimensions, by adding a 1 x 1 convolution right before the 3 x 3 and 5 x 5 filters mentioned above. Version 3 of Inception (Szegedy *et al.*, 2016) introduced to the net a factorization into smaller convolutions. The 5 x 5 convolution applied in the earlier versions is replaced by two convolutions with dimensions 3 x 3 (in a block known as Block A), reducing the number of parameters. Version 3 also brought a factorization into asymmetric convolutions, replacing the original 3 x 3 convolution into 1 x 3 and 3 x 1. Moreover, one of the new 3 x 3 convolutions is replaced by a 3 x 1 and 1 x 3 in order to reduce the parameters as well, building a new block (Block C). Block B also has a factorization, with a 1 x 7 followed by a 7 x 1 in parallel with two times a convolution of 7 x 1 and a 1 x 7. The architecture also works with reduction blocks right before their concatenations. Figure 4 shows the structure of the blocks A, B, and C, and the reduction A and B, where *n* in block B is set with value 7.

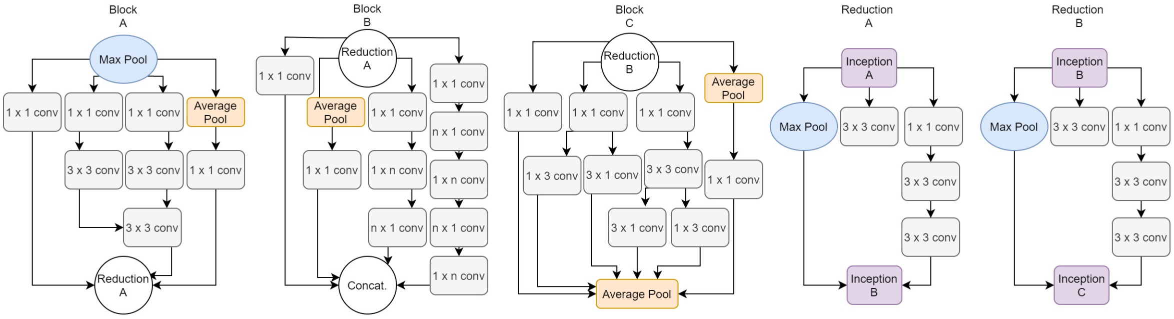

Figure 4. Inception blocks

The entire architecture of Inception-v3 is shown in Fig. 5. In total, this version has 48 layers and the architecture is also known by its intermediate classifier that attacks vanishing gradient issues, by using the calculated loss in the main classification loss, improving the results.

As reported by Szegedy *et al.* (2016), Inception-v3 was evaluated on the ILSVRC-2012 validation set, and reached a 21.2% top-1 error and 5.6% top-5 error, achieving a state of the art result.

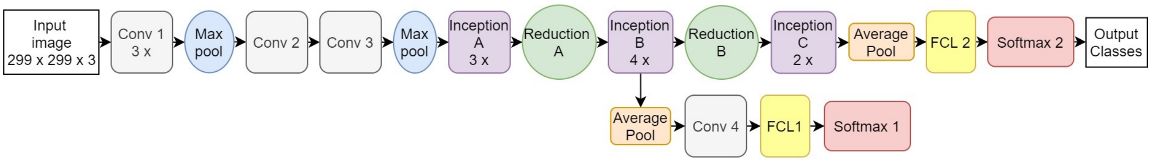

Figure 5. Inception-V3 architecture.

4.4 ResNet

Residual networks (ResNets) are a type of CNN architecture that has the ability to learn residual functions referenced to layer inputs. Before these networks, it was believed that the good results obtained by deep neural networks happened due to the additional layers inserted and increasing in the number of layers, it would be possible to obtain progressive learning in more complex problems, where each layer could extract different features. However, as shown by He *et al.* (2016a), there is empirically a depth limit for traditional CNN models. Consequently, a new type of neural network layer was introduced, the so-called residual block, which allowed for training and better performance of deep neural networks. Thus, an identity mapping is performed, a kind of bypass, which inserts the output from the previous layer to the next layer, as shown in Fig. 6.

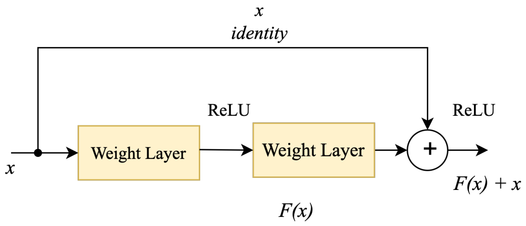

Figure 6. Residual block.

Input *x* is combined with output *F(x)* of the layer. As *x* and *F(x)* may not have the same dimensions due to the convolution operation, it is necessary to modify the identity mapping through a *W* weight function. This function adjusts the input channels to a value corresponding to the residual dimension, in order to allow the combination of these values (He *et al.*, 2016b), as shown in Eq. (2) and Eq. (3) . In the equations, x_i indicates the input and x_{i+1} the output of the i-th layer, and *F* represents the residual function. The identity mapping is represented by $h(x_i)$, which is equal to x_i and f is a ReLU function (Nair and Hinton, 2010).

$$y_i = h(x_i) + F(x_i, W_i) \tag{2}$$

$$x_{i+1} = f(y_i) \tag{3}$$

In addition, Resnet makes use of stochastic gradient descent (SGD) instead of adaptive learning techniques (Keskar and Socher, 2017) and introduced pre-processing step on input, dividing the data into patches and then feeding the network. Its main advantage is in the stacking of residual layers, and that trained achieve a good performance, unlike sequential networks where performance is reduced with increasing layers. Activation of any unit can be written as the sum of input activation with residual function. As a result, gradients can be propagated directly to shallower units, making the optimization of ResNets easier than the original mapping function and more efficient to train than very deep networks (Gu *et al.*, 2018). In this research, a ResNet18 was used due to the small size of the dataset and a few numbers of classes (binary problem).

4.5 SqueezeNet

A smaller CNN architecture with 50 times fewer parameters in comparison with AlexNet was proposed by Iandola *et al.* (2016). There are three main advantages of a CNN with fewer parameters: faster training time due to less communication across servers, more frequent new model updates to clients, and a feasible FPGA and embedded deployment. The authors focused on building an architecture called SqueezeNet that delivered these three advantages with the same level of accuracy achieved by AlexNet on ImageNet. Moreover, with a deep compression approach, SqueezeNet reached 510 times fewer parameters in comparison with AlexNet. Iandola *et al.* (2016) proposed three strategies. The first is to use 1 x 1 filters instead of 3 x 3, which reduces 9 times the number of parameters. The second strategy consists of limiting the number of input channels to 3 x 3 filters. Finally, strategy three is a process of building larger activation maps by a late down-sample, leading to higher accuracy. A fire module, consisting of a squeeze convolution layer (1 x 1 filters) followed by an expand layer (1 x 1 and 3 x 3 filters) is proposed. There are three hyperparameters in fire modules, being the number of 1 x 1 filters in the squeeze layer (s_{1x1}), the number of 1 x 1 (e_{1x1}), and 3 x 3 (e_{3x3}) filters in expanding layer. The 1 x 1 filters are responsible to implement strategy 1. In the architecture, the s_{1x1} is set to be less than e_{1x1} + e_{3x3}, limiting the number of input channels as explained in strategy 2. A fire module is illustrated in Fig. 7.

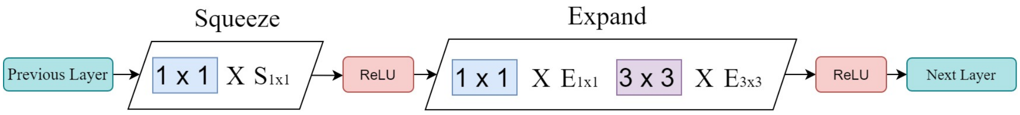

Figure 7. SqueezeNet fire block.

Three SqueezeNet architectures were constructed (Iandola *et al.*, 2016). A regular SqueezeNet, a SqueezeNet with simple bypass (skip connections), and a complex bypass SqueezeNet (skip connections with 1 x 1 convolutions). Figure 8 shows the proposed architectures. The network starts with a convolution layer (7 x 7 filter with stride 2), followed by eight fire modules. The number of hyperparameters for each fire module (s_{1x1}, e_{1x1}, and e_{3x3}) are illustrated in the same sequence. A new convolution layer (1 x 1 filter with stride 1) is then applied before the global average pool (13 x 13 filter with stride 1). The sequence ends with a soft-max function that gives the output class. The late max pooling operations (3 x 3 filter with stride 2) bring to the model larger activation maps, helping to implement strategy 3. There are no fully connected layers on these architectures, making them fully convolutional networks.

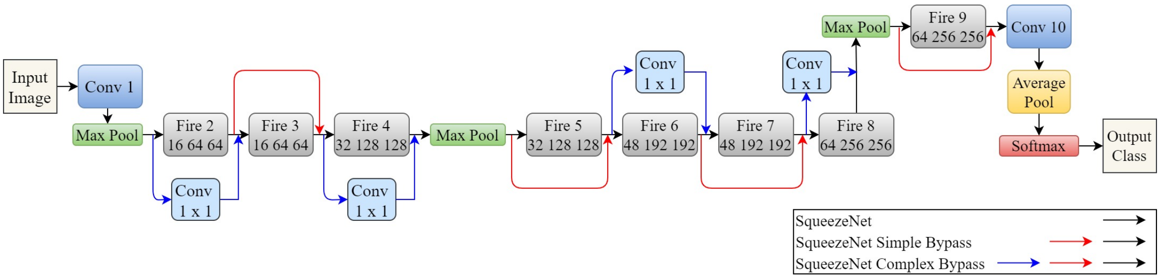

Figure 8. SqueezeNet architectures.

The architectures with bypass connections are inspired by residual networks (He *et al.*, 2016a). The simple bypass connections are element-wise additions while complex bypass with 1 x 1 convolution adds more parameters to the net. The skip connections help training the full model and improving the accuracy as well as alleviate the representational bottleneck. SqueezeNet was trained on the ImageNet ILSVRC-2012 dataset and the simple bypass architecture achieved the highest evaluation with 60.4% top-1 and 82.5% top-5 accuracy with a model size of only 4.8 megabytes.

4.6 DenseNet

The Dense Convolutional Network (DenseNet) was proposedy by Huang *et al.* (2017) and differs from other CNNs due to its feed-forward fashion way that connects each layer to all layers in the model. DenseNet has dense blocks that connects every feature map from all preceding layers to every subsequent layers. It is similar to the skip connections in Residual Networks (He *et al.*, 2016a), however DenseNet behaves very different. Due to its all connected layers, this architecture can take advantage of feature reuse, leading to a more compacted model and an implicit deep supervision. The authors (Huang *et al.*, 2017) proposed four architectures. Each one has four dense blocks with different depth. The simpler one, DenseNet-121 is shown in Figure 9. The architecture starts with a convolution and a max pool operation, and each layer in between dense blocks are known as transition layers. This version uses a fully connected layer with a thousand neurons with a softmax in order to classify the outputs.

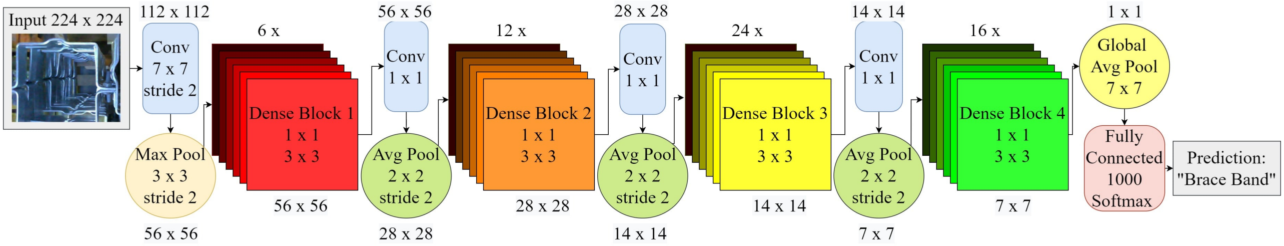

Figure 9. DenseNet-121 architecture.

DenseNets were trained on several datasets, achieving state of the art performances with the use of less computation and fewer parameters. These models tend to improve in accuracy as their dense blocks grow, showing no signs of degradation in performance or overfitting (Huang *et al.*, 2017).

5. RESULTS AND DISCUSSION

A procedure was set in order to train the models. The RGB dataset consisted of 398 images for training and 87 images for validation. Datasets described in Section 3 have a higher number of images if compared to the dataset of this paper, by the rate of: 11.54 times (Zhang *et al.*, 2021); 1.65 times (Yao and Cheng, 2021); 3.23 times (Lile and Yiqun, 2017); 2.94 times (Patel and Chowdhury, 2020); 1.58 times (Xia *et al.*, 2018). For tuning the models, it was used the same following inputs: five epochs for training, batch size equals to eight, k-fold cross-validation technique with 10 folds, cross-entropy loss function, adaptive moment estimation (ADAM) optimizer, and a learning rate of 0.001. After the training procedure was done, the loss function and accuracy of the models were compared. Figure 10 shows the models' loss (left) and accuracy (right), both for each training fold step. SqueezeNet and AlexNet achieved smaller loss values at the entire training in comparison with other models. AlexNet, SqueezeNet, and Resnet-50 had 100% accuracy on training at least in one fold, while Inception-V3 varied its accuracy and had the highest value in the loss function.

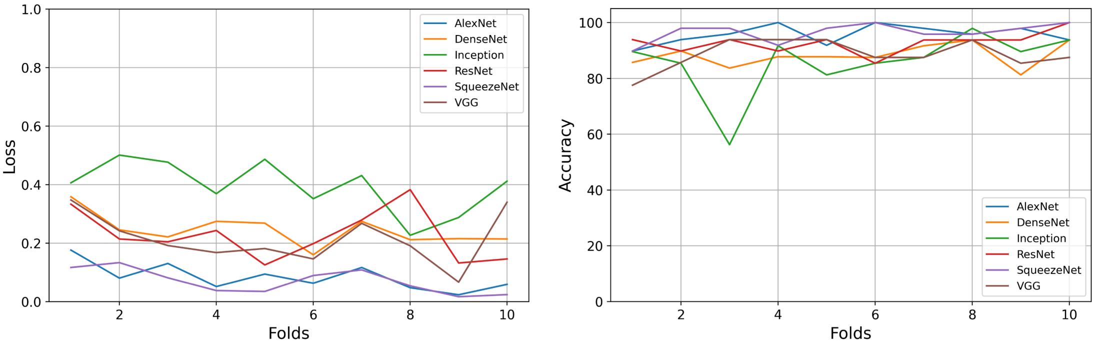

Figure 10. Training set results: models' loss (left) and accuracy (right).

The dataset used for testing consisted of 80 RGB images, being 42 samples of insulators and 38 samples of brace bands. The testing procedure intended to check the capacity of the models to deal with new data, off of the dataset used for training and validating them. This test does not intend to deal with images with different classes, images with no insulators or no brace bands. Several metrics were compared in order to check the best model to be used in this application. The chosen metrics were accuracy, precision, recall, and F1 score (Gu *et al.*, 2009). Table 1 shows the values of the metrics achieved in each model. The values in bold highlighted the best performances.

Table 1. Metrics of the evaluated models.

Models	AlexNet	VGGNet	Inception-V3	Resnet-50	SqueezeNet	DenseNet-121
Accuracy	0.938	0.913	0.888	0.938	**0.975**	0.925
Precision	0.902	0.897	0.892	0.902	**0.950**	0.864
Recall	0.974	0.921	0.868	0.974	**1.00**	**1.00**
F1 score	0.937	0.909	0.880	0.937	**0.974**	0.927

For accuracy, SqueezeNet achieved the top of the rank, with 97.5%, followed by AlexNet and Resnet-50. The accuracy measured how well the models predicted the true samples. Considering the brace band as the true positive, the precision metric measured how much the model was right in classifying it. SqueezeNet stood out with a 95% of precision, showing its ability to not classify as positive a negative sample. All models performed better with recall in comparison with their other metrics, with the exception of Inception-V3. DenseNet and SqueezeNet achieved 100% on recall, meaning that these models could classify all true positives in the testing set. F1 score shows a clearer view of the models as a whole, considering it is the harmonic mean of precision and recall. Once more, SqueezeNet achieved the best performance on the F1 score, with 97.4%. F1 score and accuracy were considered the most important metrics in this paper since it is working with a balanced dataset. SqueezeNet prevailed on these categories.

The confusion matrix in Fig. 11 shows the results achieved by SqueezeNet. The diagonal illustrates the true samples that the model predicted right (in a darker blue) while the samples off of the diagonal (in lighter blue) illustrate error in prediction. As reported by the metric recall, the model predicted all brace bands. SqueezeNet misclassed only 2 insulators. This class represents a more difficult challenge since its dataset is composed of some images with occlusion (insulator inside wooden boxes) and absence of light.

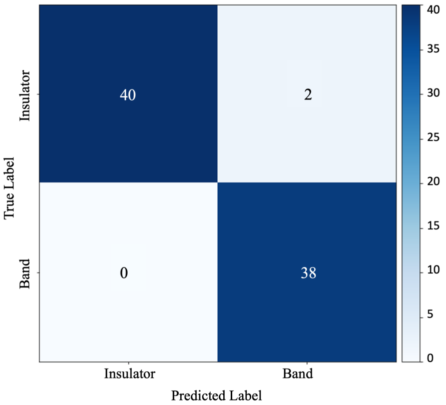

Figure 11. Confusion matrix of SqueezeNet classification.

6. CONCLUSION AND FUTURE WORKS

This paper presented a comparison of well-known CNN models applied in an inventory management classification assignment on the electric maintenance field. Six architectures were compared, being AlexNet, VGGNet-11, Inception-V3, Resnet-50, SqueezeNet, and DenseNet-121. The transfer learning method feature extraction was applied in each model, adapting the CNN for this present task by utilizing previous knowledge. Loss and accuracy metrics were observed in the training step. Afterward the training step, all models were tested with new data. Accuracy, precision, recall, and F1 score were used as metrics for evaluation. SqueezeNet stood out, achieving the best performance for this application, in terms of loss and accuracy in training as well as in all metrics on the test set. The results are consistent with the values found in the literature for this type of classification problem, given the particularities of each dataset.

For future studies, it is intended to experiment with an ensemble of classifiers, as well as tuning them individually, in order to improve performance and deliver more reliability for an intelligent inventory management solution. An expansion of the dataset is also needed, for a better representation and description of such tasks.

7. ACKNOWLEDGEMENTS

The authors thank the financial support of the Agência Nacional de Energia Elétrica (ANEEL) and Companhia Paranaense de Energia (COPEL DIS), by the Project n° PD-02866-0011/2019, and Instituto de Tecnologia para o Desenvolvimento (LACTEC) for the assistance.

8. REFERENCES

Al-Momani, H., Al Meanazel, O.T., Kwaldeh, E., Alaween, A., Khasaleh, A. and Qamar, A., 2020. "The efficiency of using a tailored inventory management system in the military aviation industry". *Heliyon*, Vol. 6, No. 7, p. e04424.

Chen, Z. and Dong, R., 2021. "Research on fast recognition method of complex sorting images based on deep learning". *International Journal of Pattern Recognition and Artificial Intelligence*, p. 2152005.

Fan, K., Niu, L. and Zhang, S., 2020. "E-commerce item identification based on improved squeezenet". In *Journal of Physics: Conference Series*. IOP Publishing, Vol. 1626, p. 012002.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. *et al.*, 2018. "Recent advances in convolutional neural networks". *Pattern Recognition*, Vol. 77, pp. 354–377.

Gu, Q., Zhu, L. and Cai, Z., 2009. "Evaluation measures of the classification performance of imbalanced data sets". In *International symposium on intelligence computation and applications*. Springer, pp. 461–471.

He, K., Zhang, X., Ren, S. and Sun, J., 2016a. "Deep residual learning for image recognition". In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778.

He, K., Zhang, X., Ren, S. and Sun, J., 2016b. "Identity mappings in deep residual networks". In *European conference on computer vision*. Springer, pp. 630–645.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. "Densely connected convolutional networks". In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4700–4708.

Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J. and Keutzer, K., 2016. "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size". *arXiv preprint arXiv:1602.07360*.

Keskar, N.S. and Socher, R., 2017. "Improving generalization performance by switching from adam to sgd". *arXiv preprint arXiv:1712.07628*.

Khan, A., Sohail, A., Zahoora, U. and Qureshi, A.S., 2020. "A survey of the recent architectures of deep convolutional neural networks". *Artificial Intelligence Review*, Vol. 53, No. 8, pp. 5455–5516.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. "Imagenet classification with deep convolutional neural networks". *Advances in neural information processing systems*, Vol. 25, pp. 1097–1105.

LeCun, Y., Bengio, Y. *et al.*, 1995. "Convolutional networks for images, speech, and time series". *The handbook of brain theory and neural networks*, Vol. 3361, No. 10, p. 1995.

Lichtenthaler, U., 2020. "Beyond artificial intelligence: why companies need to go the extra step". *Journal of Business Strategy*, Vol. 41, No. 1, pp. 19–26. doi:10.1108/jbs-05-2018-0086.

Lile, C. and Yiqun, L., 2017. "Anomaly detection in thermal images using deep neural networks". In *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 2299–2303.

Nair, V. and Hinton, G.E., 2010. "Rectified linear units improve restricted boltzmann machines". In *Icml*.

Patel, A.D. and Chowdhury, A.R., 2020. "Vision-based object classification using deep learning for inventory tracking in automated warehouse environment". In *2020 20th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, pp. 145–150.

Preet Kour, V. and Arora, S., 2020. "Vision based techniques for image classification: a survey". *Sakshi, vision based techniques for image classification: a survey (March 28, 2020)*.

Purdy, M. and Daugherty, P., 2017. "How ai boosts industry profits and innovation (web publications)". France, https://www.accenture.com/fr-fr/_acnmedia/36dc7f76eab444cab6a7f44017cc3997.pdf. Accessed on: Feb. 6, 2021.

PwC, 2017. "Sizing the prize: what's the real value of ai for your business and how can you capitalise". USA, https://www.pwc.com/gx/en/news-room/docs/report-pwc-ai-analysis-sizing-the-prize.pdf. Accessed on: Feb. 6, 2021.

Relyea, R., Bhanushali, D., Manghi, K., Vashist, A., Hochgraf, C., Ganguly, A., Kwasinski, A., Kuhl, M.E. and Ptucha, R., 2020. "Improving multimodal localization through self-supervision". *Electronic Imaging*, Vol. 2020, No. 6, pp. 14–1.

Sanghvi, K., Aralkar, A., Sanghvi, S. and Saha, I., 2020. "A survey on image classification techniques". *Available at SSRN 3754116*.

Simonyan, K. and Zisserman, A., 2014. "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv:1409.1556*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. "Going deeper with convolutions". In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. "Rethinking the inception architecture for computer vision". In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2818–2826.

Tejesh, B.S. and Neeraja, S., 2018. "Warehouse inventory management system using iot and open source framework". *Alexandria Engineering Journal*, Vol. 57, No. 4, p. 3817–3823. doi:10.1016/j.aej.2018.02.003.

Tian, G., Meng, S., Bai, X., Zhi, Y., Ou, W., Fei, X., Tan, Y. *et al.*, 2020. "Electric tower target identification based on high-resolution sar image and deep learning". In *Journal of Physics: Conference Series*. IOP Publishing, Vol. 1453, p. 012117.

Torrey, L. and Shavlik, J., 2010. "Transfer learning". In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, IGI global, pp. 242–264.

Wamba-Taguimdje, S.L., Fosso Wamba, S., Kala Kamdjoug, J.R. and Tchatchouang Wanko, C.E., 2020. "Influence of artificial intelligence (ai) on firm performance: the business value of ai-based transformation projects". *Business Process Management Journal*, Vol. 26, No. 7, pp. 1893–1924. doi:10.1108/bpmj-10-2019-0411.

Xia, Y., Lu, J., Li, H. and Xu, H., 2018. "A deep learning based image recognition and processing model for electric equipment inspection". In *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*. IEEE, pp. 1–6.

Yang, J.X., Li, L.D. and Rasul, M.G., 2021. "Warehouse management models using artificial intelligence technology with application at receiving stage-a review". *International Journal of Machine Learning and Computing*, Vol. 11, No. 3, pp. 242–249.

Yao, N. and Cheng, K., 2021. "Electric power equipment image recognition based on deep forest learning model with few samples". In *Journal of Physics: Conference Series*. IOP Publishing, Vol. 1732, p. 012025.

Zhang, J., Lu, C., Li, X., Kim, H.J. and Wang, J., 2019. "A full convolutional network based on densenet for remote sensing scene classification". *Math. Biosci. Eng*, Vol. 16, No. 5, pp. 3345–3367.

Zhang, Q., Chang, X., Meng, Z. and Li, Y., 2021. "Equipment detection and recognition in electric power room based on faster r-cnn". *Procedia Computer Science*, Vol. 183, pp. 324–330.

Zhang, S., Wu, Y. and Chang, J., 2020. "Survey of image recognition algorithms". In *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, Vol. 1, pp. 542–548.

9. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.