A Comparative Study of Methods based on Deep Neural Networks for Self-reading of Energy Consumption in a Chatbot Application Context

Carlos V. M. Rocha*, Pedro H. C. Vieira*, Antonio M. Pinto*, Pedro V. Bernhard*,
Ricardo J. F. Anchieta Junior*, Ricardo C. S. Marques*, Italo F. S. Silva*,
Simara V. Rocha*, Aristófanes C. Silva*, Hugo D. C. S. Nogueira†, Eliana M. G. Monteiro†
*Applied Computing Group (NCA), Federal University of Maranhão (UFMA), São Luís - MA, Brazil
Email: {carlos.martins, pedro.carvalho, antoniomp, pedro.bernhard,
ricardo.anchieta, ricardo.marques, francyles, simara, ari}@nca.ufma.br
†Equatorial Energy Group, Brazil
Email: {hugo.nogueira, eliana.monteiro}@equatorialenergia.com.br

***Abstract*—Self-reading is a process in which the consumer is responsible for measuring his own energy consumption, which can be done through digital platforms, such as websites or mobile applications. The Equatorial Energy group's electric utilities have been working on developing a chatbot application through which consumers can send an image of their energy meter to a server that runs a method based on image processing and deep learning for the automatic recognition of consumption reading. However, the incorporation of these methods in a solution available to the public should consider factors such as response time and accuracy, so that it presents a satisfactory response time when it needs to handle a large number of simultaneous requests. Therefore, this paper presents a comparative study between approaches developed for the automatic recognition of consumption readings in images of electric meters sent to the server. Response time performances are analyzed through stress tests that simulate the real application scenario. The mean average precision (mAP) and the accuracy metrics of the methods are also analyzed in order to evaluate the generalization of the used convolutional neural networks.**

I. Introduction

The Equatorial Energy group's electric utilities measure the energy consumption with the help of the meter readers that are the employees responsible for collecting information from the consumers' energy meters. The meter readers manually record readings on a mobile device that processes this information and prints an invoice with the amount of monthly consumption. However, because it is done manually, this process is susceptible to errors, which can lead to problems in the reading and billing processes. These problems are called non-technical losses by the Brazilian Electricity Regulatory Agency (ANEEL) [1].

The use of smart energy meters emerges as a possibility for those problems to be mitigated. However, this solution has a high financial cost and demands an implementation time that is impracticable in the short term, due to the number of meters to be replaced. A more viable alternative is the self-reading, which comprises the use of digital platforms, such as websites or mobile applications through which the consumer himself can perform your reading consumption. It is important to notice that this practice has been encouraged by ANEEL through a normative resolution[1].

Self-reading creates a closer relationship between utilities and consumers, as the latter become part of the consumption reading process. This aims to reduce the occurrence of errors and fraud, especially in areas of difficult access and inspection, such as rural areas. In addition, regarding the meter readers, it is avoided that they are exposed to different weather conditions for long hours of the day to take the readings, a factor that can cause them fatigue and health problems.

Self-reading is strongly related to the use of digital platforms. Thus, the development of computational methods to automate this process can make its execution easier, in addition to bringing speed and security. In this context, image processing techniques and computational intelligence are possible alternatives to be explored. In this case, the reading process would have as input an image of a energy meter, instead of the numerical sequence of reading that could be interpreted wrongly by the consumer.

The Equatorial Energy group has a chatbot application integrated with Whatsapp[2] though which various services are offered, such as invoice emission and requesting repairs to the electrical network. This application is used by a significant part of consumers in the utilities' coverage areas. This encouraged the development of a self-read service to be integrated with this chatbot application. And as the operation of this application is well known by consumers, the self-reading process can be understood more easily.

However, to incorporate self-reading computational methods into a solution available to the pubic, it is important to consider factors such as response time and accuracy, so that the automatic reading process obtains accurate results in a short time. Based on that, this paper presents a comparative study

[1]Available on http://www2.aneel.gov.br/cedoc/ren2020878.pdf
[2]Available on https://www.whatsapp.com/

between image recognition methods in the context of self-reading as a service to be incorporated into a chatbot solution. In this scenario, the consumer initiates a conversation with the virtual assistant. At a certain point, the consumer is asked to send an image of his energy meter. This image is sent to a server where it will be processed by a method based on image processing and convolutional neural networks for automatic recognition of the meter reading digits.

The study is based on the hypothesis that it is possible to develop a method whose response time is reduced and the accuracy performance is not degraded. Therefore, this paper presents a comparative analysis of the performances achieved by approaches developed in relation to time and accuracy factors.

This paper is organized as follows: Section II presents the related works. Section III describes the development of the comparative study; the obtained results are reported in Section IV; and, finally, the conclusion and future works are presented in Section V.

II. Related Works

In the context of automatic reading of energy consumption, there are works in the literature that present solutions based on image processing, traditional machine learning techniques and deep learning. These works aim at recognizing digits displayed by meters.

The method proposed in [2] comprises a hybrid approach for digit recognition in meters. MobileNet V2 [3] is used for location of the *display* region. This region is then binarized and passed as input to the digit segmentation. The segmented digits are classified via the Support Vector Machine (SVM) [4]. The experiments were performed on a private dataset and the method achieved 88.67% accuracy in a test with 300 meters images.

The method presented in [5] is based on the joint use of specific convolutional neural networks (CNN) to detect the meter's display and predict the length of the reading sequence, as well as its value. The experiments were carried out with a private dataset containing images of different meter models (gas, water and electricity), resulting in an accuracy per digit of 85.70%.

The mentioned methods have many parameters and thresholds to be adjusted based on image sets with a small number of examples. These factors limit the reproducibility of these approaches for the context of self-reading, which is characterized by different lighting conditions for image capture, and also by the high variability of meter models.

In [6], experiments with YOLO [7] and Faster R-CNN [8] networks are presented for reading recognition in images from the UFPR-ARM public database [9], which is composed of analogical and digital meters. With Faster R-CNN, the authors obtained 93.60% as the best result of overall accuracy per digit. In [10], experiments on this dataset are also conducted in order to validate a method based on Mask-RCNN with GoogLeNet backbone [11]. This approach achieved an overall accuracy of 99.86% for digit recognition.

These approaches are promising in the recognition of reading digits. However, they were not designed taking into account the context of a self-reading application, in which response time is also an important factor to be evaluated.

In [12], digit recognition in images of analogical and digital meters is addressed without considering previous steps of detection and segmentation. Features of each digit image are extracted via Histogram of Oriented Gradients (HoG) and Local Self-Similarity (LSS). The results obtained for accuracy per digit were 96.72% and 97.90%, respectively, for analogical and digital meters. In [13], it is presented a method for the detection and segmentation of meters, displays and identification components as a prior step to recognition and also as a way to ensure safety in the reading process. Single Shot Multibox Detector network [14] were used to detect the objects of interest, obtaining 73.10% of mean average precision (mAP) for all generated bounding boxes.

The methods mentioned were designed to be incorporated into a mobile application, and their execution requires processing and storage capacity of the devices. In [15], a method for self-reading is presented as a service incorporated into a chatbot solution. The digit recognition method is based on the combined use of RetinaNet [16] and an ensemble composed of SVM, Xgboost [17] and Efficient Net [18] classifiers. This approach achieves overall accuracy of 89% and 77.20%, respectively, for analogical and digital meters. It is noteworthy that this method is performed on the server side, which is more prepared compared to the mobile device.

The present comparative study is based on [15]. In this work, limitations are identified and analysed. In addition, improvements are proposed based on a comparative analysis between other methods for the automatic recognition of meter readings within the scope of a chatbot solution for self-reading of energy consumption. Differently from previous works, Scaled Yolo v4 [19] and EfficientDet [20] networks are used in this work. These approaches present mechanisms to improve the learning process and reduce the inference time.

III. Materials and Methods

In this section will be presented the proposed method for energy consumption self-reading and the techniques used during the preparation of this comparative analysis jointly with a description of the image dataset used in the experiments.

A. Dataset

The used dataset contains digital and analogical energy meters images, captured by meter readers in different illumination conditions, positions and using various mobile device types. Meters usually are involved by a transparent protection case which allows the visualization of the consumption digit sequence. Some examples can be seen in Figure 1.

The dataset contains 8105 meter images, of which 4226 belong to analogical class and 3879 to digital class. Among the images captured there are some cases in which the meter‘s protective case presents some damage due to weathering or human action, which makes it difficult to see the meter and

Fig. 1. Examples of energy meter images present in dataset.

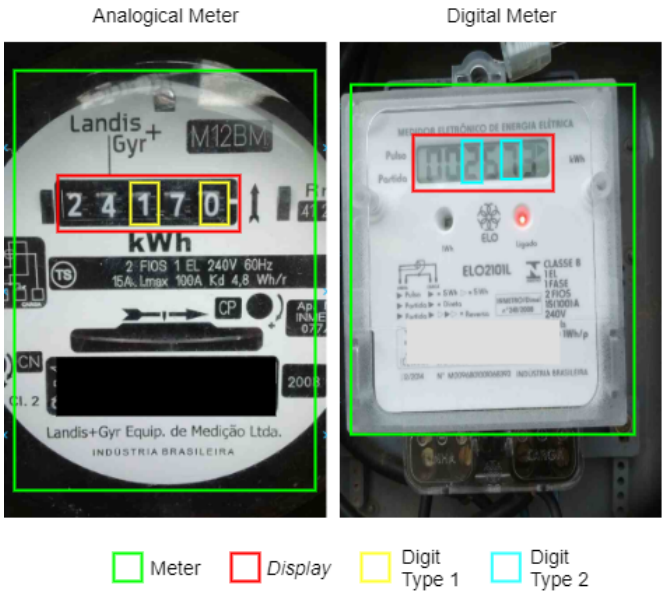

Fig. 2. Examples of dataset images and annotations.

other components. However, it was decided not to discard these images due the necessity of developing a robust method that could be applied in the real world.

Fig. 3. Examples of digits founded in the dataset.

Each image in the dataset is accompanied by an .xml file that contains the bounding box coordinates of the meter, display and the digits. Figure 2 shows an example of those annotations. In the case of the digits, they are divided between types 1 and 2, referring to analogical and digital meters respectively.

In the case of the analogical meters, the digits may vary according to the model and manufacturer, as illustrated in Figure 3. In relation to digital meters, they have a standardized format. Finally, besides receiving the labels aforementioned, the digits of type 1 and 2 also have complementary information referring to their classification between 0 and 9, which is important for the training of classification models.

B. Proposed Method

The proposed method for energy consumption reading recognition consists of two main steps: initial detection and reading recognition. An overview of these steps can be seen in Figure 4.

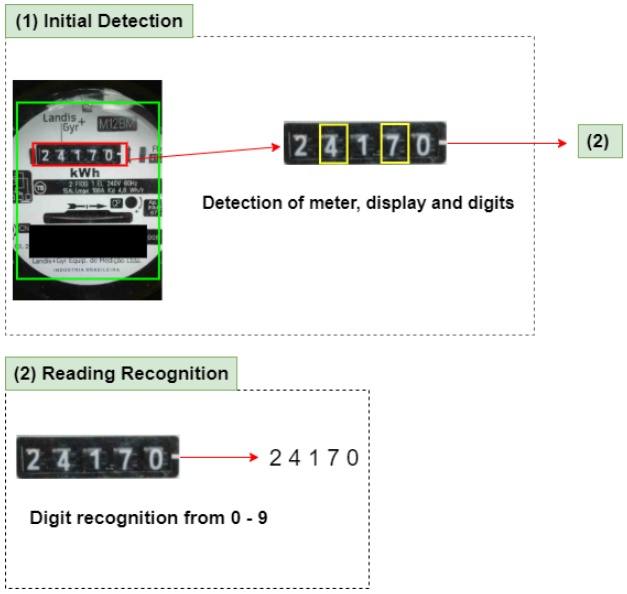

Fig. 4. Overview of the proposed method for recognizing energy consumption readings.

The consumer send an image via chatbot to the server (called inference server) that runs the method. In the initial detection step, it will first identified whether the meter is present in the image. If this is confirmed, the method goes on to detect the display and the reading digits. Otherwise, a message is sent to the consumer saying that the meter has been not found. Thus, it is possible to avoid errors in the reading process whether the consumer mistakenly or deliberately sends an image without a meter.

Then, the detected digits serve as input for the reading recognition step, where each detected digit will be classified between numbers 0 to 9. Finally, the server will send the complete sequence of digits recognized to the chatbot, which will show this result to the consumer to proceed with the self-reading process.

The method proposed in [15], that is the main related work, processes the images in a specific way for each type of meter (analogical or digital). The type of processing to be performed depends on the initial detection result. In this step two RetinaNet networks are used, each trained to identify one type of meter. Thus, an image with dimensions 1280x960 is initially submitted to these two networks, and the one obtaining results with the highest confidence rate will define which processing should be performed. For the detection of

the display digits, two RetinaNet networks are also used, one for type 1 digit location (analogical) and another for type 2 (digital). The network to be run is chosen according to the result obtained in the meter detection. Finally, each located digit is classified through an ensemble composed by EfficientNet, SVM and Xgboost classifiers.

As can be seen, the are a large number of models used in the mentioned method. This leads to increased execution time, since many models will need to be loaded in memory to attend each request made to the server. Therefore, in a scenario with multiple requests, the response time will be increased. Due to these limitations, the present work attempted to analyze other techniques in order to achieve promising results, optimize the use of server resources, and provide results in a more agile way.

Thus, this work proposes, for the initial detection step, the use of a single convolutional neural network (CNN) to process meter images regardless of their type (analogical or digital). The objects of interest should be the meter, display, and digits (types 1 and 2), and this process should occur in a single step (one-shot). In this way, the loading of several models, each focused on a specific type of meter and digits, is avoided. Following this guideline, experiments were conducted with the Scaled-Yolo v4 and EfficientDet networks, which have in their architectures some mechanisms designed to reduce the inference time without degrading the accuracy of the models.

The Yolo v4 [21] architecture presents a mechanism called Cross Spatial Partial (CSP), initially introduced by [22] which helps in reducing the computational cost provided by the high amount of inference computations performed during network training, which is a problem faced by other architectures, such as ResNet [23] or its predecessor, Yolo v3 [24]. Scaled-Yolo v4 [19] is an approach focused on improving Yolo v4 in order to balance the factors of inference time and accuracy in the scope of object detection. It was developed from the joint scaling between the factors of input size and number of stages (aiming at better accuracy); width and depth, according to the requirements of inference time. The architecture used in this work is P5, as it is a simpler model but presents worthy results for the MSCOCO 2017 database [3].

The other architecture used in this work is the EfficientDet D1 [20]. The models of the EfficienDet family (D0 to D7) are pioneers regarding scaling based on the optimization of scaling thresholds, considering also the size of the input images, the width and depth of the network. The larger these factors are, the more complex the model is, thus leading to a high inference time, which is something to be avoided in the application scenario of the proposed method. Therefore, based on this, the D1 architecture was chosen. It should be noted that, in this work, Scale Yolo V4 and EfficientDet D1 receive inputs with dimensions 640x640 defined from tests performed.

At the end, the bounding boxes generated in the initial detection step are used to extract the isolated digit cropping them from the images. Each digit crop is resized to 80x80 (with three channels) and then passed to the EfficientNet B2 that is responsible for the classification between 0 and 9.The choice of this network is based on the promising results obtained in the ImageNet and ImageNet-V2 [25] challenges, both involving image classification. That network is composed of depth convolution blocks, which have fewer parameters to be optimized compared to traditional convolution [26]. These blocks act in conjunction with the Squeeze-and-Excitation mechanism [27], whose function is to assign weights to the feature maps, so that those with greater relevance get more weight than others. This way, EfficientNet B2 can perform a more efficient feature extraction without significantly increasing the computational cost.

The proposal of a single model for performing the digit recognition task after initial detection aims to avoid having many models loaded into memory as a result of specialized training for each type of digit (type 1 and 2).

After recognizing the digits of the reading, the method returns this result to the chatbot, that will show to the consumer the value of the reading. Thus, he can make some correction in case of failure or continue the process if the reading is correct.

IV. Experiments

In this Section, the results obtained by the study carried out in this work are presented. It is also shown more details about the stress testing protocol of the developed approaches and the evaluation metrics used.

A. Stress Testing Protocol and Metrics

To carry out the stress tests, the open source library Locust[4] was used, implemented in the language Python[5]. With this library, multiple requests to an endpoint on the web are simulated. In this case, the endpoint is the inference server, which is a web service. Each request represents a user's access, so tests were performed with different amounts of simultaneous accesses in order to evaluate the response time of each approach developed for reading recognition. These quantities are defined as powers of base 2, thus ranging from 2 to 1024 simultaneous requests.

In the tests, the average response time of the requests was collected, in addition to the minimum and maximum times, the latter being related to the fulfillment of the last request. Thus, it is possible to evaluate the performance of the methods in terms of speed. Regarding the evaluation of the networks' performances, the metrics used are mean average precision (mAP) and accuracy per reading sequence. The first is commonly used to validate object detection methods; and the second aims to evaluate the process of recognizing the reading digits.

B. Results

8105 electric meter images were divided into the following proportions: 70% for training (5629 images) and 30% for

[3]Available on https://cocodataset.org/

[4]Available on https://github.com/rednafi/stress-test-locust

[5]Available on https://www.python.org/

TABLE I
RESPONSE TIMES OBTAINED IN STRESS TESTS. THE AVERAGE (AVG), MAXIMUM (MAX.) AND MINIMUM (MIN.) USER SERVICE TIMES WERE COLLECTED.

	Response Times (in seconds)								
	[15]			Scaled Yolo v4 + EffB2			EfficientDet D1 + EffB2		
Users	Avg.	Min.	Max.	Avg.	Min.	Max.	Avg.	Min.	Max.
2	2.8	2.5	3.0	**0.8**	**0.8**	**0.8**	**0.8**	**0.8**	**0.8**
4	5.9	1.5	7.6	**0.8**	**0.8**	**0.8**	**0.8**	**0.8**	**0.8**
8	11.5	5.8	14.3	1.1	**0.8**	1.4	**0.8**	**0.8**	**0.8**
16	8.5	3.8	13.3	5.4	**0.8**	8.3	3.2	**0.8**	5.1
32	14.9	3.0	26.0	12.7	**0.8**	15.0	4.7	**0.8**	7.2
64	28.6	4.2	53.0	12.5	1.0	19.0	10.1	1.0	13.0
128	55.2	3.4	106.4	17.5	1.9	30.6	13.5	**0.8**	23.0
256	108.4	3.3	213.4	34.6	2.8	67.5	21.1	1.4	41.2
512	217.6	4.4	432.0	57.7	1.2	120.6	36.8	1.0	73.0
1024	370.6	4.7	895.0	**119.3**	1.6	247.0	**65.0**	1.3	128.0

TABLE II
MAP RESULTS OBTAINED BY EACH TESTED APPROACH FOR INITIAL DETECTION. DM AND AM CORRESPOND, RESPECTIVELY, TO DIGITAL AND ANALOGICAL METERS.

	mAP (%)		
	[15]	EfficientDet D1	Scaled Yolo v4
Display (DM)	88.0	**98.00**	**98.00**
Display (AM)	**100.00**	98.00	99.00
Digit (Type 1)	**92.00**	80.00	88.00
Digit (Type 2)	**98.00**	88.00	95.00
Digital Meter	**97.00**	93.00	92.00
analogical Meter	**97.00**	86.00	87.00
Overall	**95.33**	90.50	93.16

testing (2476 images). With a training set larger than the test set, the aim is to favor the learning process of the models. As discussed in Section III-A, the dataset is quite heterogeneous. Besides the differences between digital and analogical types, there are also intra-type divergences, as meters also vary according to manufacturers. Thus, it is necessary that the training set has a relevant number of examples of each model. As for the test, it is guaranteed an expressive number of images to simulate the real application scenario, also marked by the heterogeneity of the meters.

The convolutional networks were implemented by using Keras (version 2.4) and Tensorflow (version 2.3) [28] libraries. EfficientDet D1 was trained for 50 epochs using the Adam optimizer with a learning rate (lr) of 10^{-3}, and the loss functions Smooth L1, for regression, and Focal Loss, for classification. Scaled Yolo v4 was trained for 200 epochs. Other hyperparameters was: Adam optimizer, lr equal to $2x10^{-5}$ and the loss functions Complete Intersect over Union (CIoU) [29] and Binary Crossentropy, used respectively for regression and classification. Finally, the Efficient Net B2 was trained for 50 epochs with the RMSProp optimizer (lr of $2x10^{-5}$) and the Categorical Crossentropy loss function. To reproduce the method proposed in [15], the same hyperparameters suggested in that work were used.

In Table I, the results obtained with the stress tests are shown. As mentioned, the average, minimum and maximum times referring to the responses of simultaneous requests were collected. The minimum and maximum times refer, respectively, to the service of the first and last user. Henceforth, EffB2 will be the defined abbreviation for EfficientNet B2.

The approach of [15] takes an average of 2.8 seconds to serve two simultaneous users. In contrast, approaches that use EfficienDet and Scaled Yolo v4 in the initial detection step reaches 0.8 seconds, a reduction of approximately 72%. For the latter approaches, this average time is also maintained with four users, but starts to change when this number increases to eight concurrent users.

Overall, EfficientDet D1 had the lowest response times compared to other approaches. In the most critical scenario, with 1024 concurrent users, the average response time is 65 seconds, with the last user answered in 128 seconds. Scaled Yolo v4, in turn, gets an average time of 119.3 seconds; and the approach proposed in [15] takes 370.6 seconds, answering the last user in 895 seconds.

In Figure 5, it can be seen that EfficientDet D1 and Scaled Yolo v4 achieved significantly reduced response times in all test scenarios with different numbers of users. This shows the impact of centralizing the recognition process, because as these approaches do not use a large set of networks, so the number of input and output routines for loading models into memory is reduced.

However, besides the analysis of response times, it is necessary to consider the accuracy obtained by the recognition methods of the reading digits. For this, the mAP results related to the initial detection step, and the accuracy, were analyzed to evaluate the reading recognition step. The results of the initial detection can be seen in Table II.

Overall, the method proposed in [15] outperforms the oth-

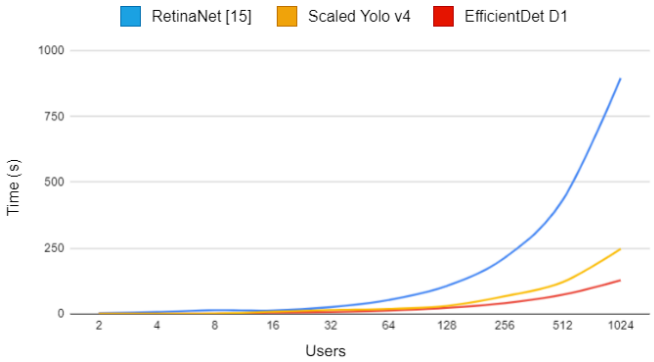

Fig. 5. Comparative analysis of response times.

ers, achieving the best mAP results for detection of digits and electric meters. Its less expressive result refers to the detection of digital meter displays. In this specific case, EfficientDet D1 and Scaled Yolo v4 achieve better mAP values. However, their performance is inferior to the other components. This is expected since in [15] specialized trained networks are used to detect each component in a specific type of meter. Differently, the training process of the other approaches involves learning features of many elements, with the side effect of increasing the chances of failure, which is reflected in the degradation of mAP. Nevertheless, the overall mAP result of Scaled Yolo v4 (93.16%) is the second best, indicating that the idea of reducing the number of networks used in the initial detection step is promising.

Finally, the results of the reading digit recognition step can be seen in Table III. It is important to note that, in this case, the accuracy metric is used to evaluate the recognition of the complete sequence of reading digits. Thus, if a digit is misclassified, the entire sequence will be considered wrong.

TABLE III
ACCURACIES OBTAINED BY EACH TESTED APPROACH FOR RECOGNITION OF READING DIGITS.

Method	Accuracy (%)
YoloV4 Scaled + EffB2	**60.00**
[15]	59.00
EfficientDet D1 + EffB2	50.00

The approach that combines EfficientDet D1 and EffB2 obtains much lower accuracy compared to the others (50%). On the other hand, Scaled Yolo v4 + EffB2 achieves 60% accuracy, matching the result obtained by [15], which is 59% for the image test set used in this study.

According to stress tests, approaches using EfficientDet D1 and Scaled Yolo v4 are faster. However, among them, only Scaled Yolo v4 showed a promising accuracy value, which was only 1% higher compared to [15]. Therefore, the initial hypothesis of this study is corroborated, because the experiments show that is possible to optimize the response time without significantly degrading the accuracy.

Thus, based on the comparative study between the mentioned approaches, it is understood that, for the inference server to attend a larger number of users in less time, the Scaled Yolo v4 + EffB2 method would be a viable alternative. However, it is realized that, despite being promising, this method can be improved to obtain better values of mAP and accuracy.

V. CONCLUSION

This work presented a comparative study between approaches based on convolutional neural networks applied to the self-reading of energy consumption through a chatbot solution developed by the Equatorial Energy group. In this solution, consumers send an image of their meter through the chatbot application to the inference server responsible for executing a method that performs automatic reading recognition. This method consists of steps: first, the initial detection of the meter, display and display digits; and then the classification of the digits. Finally, the recognized reading value is sent to the consumer.

The approaches used as the object of analysis were subjected to stress tests to evaluate the performance to response time. In addition, the mAP and accuracy metrics related to the recognition method were also evaluated. Stress tests show that EfficientDet D1 fulfills 1024 simultaneous requests with an average time of 65 seconds. For this same scenario, Scaled Yolo v4 reached an average time of 119.3 seconds. On the other hand, the method proposed in [15] takes an average of 370.6 seconds.

As for accuracy, Scaled Yolo v4 + EffB2 obtained values equivalent to [15], showing that it is possible to reduce the response time without degrading the result for automatic reading recognition, thus confirming the main hypothesis of this study. Therefore, among the analyzed approaches, Scaled Yolo v4 + EffB2 is a promising alternative to be incorporated into the inference server in the context of the chatbot solution under development. However, it is understood that this accuracy value can be improved to provide more precision to the self-reading process.

As future work, it is intended to continue the study aiming to improve the accuracy of reading recognition. Tests will be carried out with other EfficientNet models (B and L families) and also networks based on Vision Transformers [30], which have achieved meritorious results in the image classification task. Furthermore, it is intended to incorporate into the chatbot solution (under development) a method for recognizing the identification codes of the meters (called tags) as an additional step ensure more security for the self-reading process.

ACKNOWLEDGEMENTS

This work is part of the AutoClara project, financially supported by Equatorial Energy Group under the Brazilian Electricity Regulatory Agency's (ANEEL) R&D program (code: APLPED0004-PROJETOPED-0036-S01).

REFERENCES

[1] ANEEL. (2019, apr) Energia no brasil e no mundo. [Online]. Available: http://www2.aneel.gov.br/arquivos/pdf/atlas_par1_cap2.pdf

[2] H. Shuo, Y. Ximing, L. Donghang, L. Shaoli, and P. Yu, "Digital recognition of electric meter with deep learning," in *2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*. IEEE, 2019, pp. 600–607.

[3] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation," *CoRR*, vol. abs/1801.04381, 2018. [Online]. Available: http://arxiv.org/abs/1801.04381

[4] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[5] A. Calefati, I. Gallo, and S. Nawaz, "Reading meter numbers in the wild," in *2019 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2019, pp. 1–6.

[6] G. Salomon, R. Laroca, and D. Menotti, "Deep learning for image-based automatic dial meter reading: Dataset and baselines," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[7] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, 2015. [Online]. Available: http://arxiv.org/abs/1506.02640

[8] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015. [Online]. Available: http://arxiv.org/abs/1506.01497

[9] R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz, and D. Menotti, "Convolutional neural networks for automatic meter reading," *Journal of Electronic Imaging*, vol. 28, no. 1, pp. 1 – 14, 2019. [Online]. Available: https://doi.org/10.1117/1.JEI.28.1.013023

[10] A. Azeem, W. Riaz, A. Siddique, and U. A. K. Saifullah, "A robust automatic meter reading system based on mask-rcnn," in *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*. IEEE, 2020, pp. 209–213.

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

[12] A. Serra, J. França, R. Marques, W. Figueredo, A. Reis, I. Santos, S. Rocha, A. Silva, E. Monteiro, I. Silva, M. Silva, and J. Santos, "Reconhecimento de dígitos em imagens de medidores de energia no contexto de um aplicativo de autoleitura," 01 2019.

[13] A. Serra, J. França, J. Sousa, R. Costa, I. Santos, S. Rocha, A. Silva, A. Paiva, E. Monteiro, I. Silva, M. Silva, and J. Santos, "Segmentação semântica de medidores de energia elétrica e componentes de identificação," 01 2019.

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," *CoRR*, vol. abs/1512.02325, 2015. [Online]. Available: http://arxiv.org/abs/1512.02325

[15] C. V. Rocha, G. X. Bras, J. E. Oliveira, A. G. Fernandes, A. A. Lima, A. M. Paiva, I. F. S. da Silva, S. V. da Rocha, E. M. G. Monteiro, and E. C. Fernandes, "Uma solução de chatbot para a realização de autoleitura do consumo de energia por meio de aplicativos de mensagens," *Anais da Sociedade Brasileira de Automática*, vol. 2, no. 1, 2020.

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[17] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[18] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. Long Beach, California, USA: PMLR, 09–15 Jun 2019, pp. 6105–6114. [Online]. Available: http://proceedings.mlr.press/v97/tan19a.html

[19] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13 029–13 038.

[20] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 781–10 790.

[21] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[22] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 390–391.

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

[24] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[25] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar, "Do imagenet classifiers generalize to imagenet?" in *International Conference on Machine Learning*. PMLR, 2019, pp. 5389–5400.

[26] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[28] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[29] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-iou loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 993–13 000.

[30] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.