Bootcamp Final Task

Defining the problem, understanding the study area, creating a dataset, proposing new methods, fine tuning and evaluating a deep learning model.

Breno da Silva Lima

Bachelor of Science in Computer Science

Federal Technological University of Paraná

Ponta Grossa, Paraná

brenolima@vamos.utfpr.edu.br

***Abstract*—As the final work of the Machine Learning Bootcamp, this project aims to consolidate the knowledge acquired throughout the tasks into its own project, exploring the entire pipeline of a machine learning problem. In this article, a real problem was identified, from which a solution was proposed: the application of OCR to water meters to extract water consumption readings. The study area was investigated in depth, and new methods were proposed, resulting in significant improvements to the model used.**

***Key-words*—OCR, water meters, computer vision.**

I. Introduction

Currently in the city of São Paulo, water consumption measurements are carried out by Sabesp, which monitors consumption in houses and condominiums. Since 2021, as is the case for houses, readings for condominium residents must be carried out individually, that is, each resident must receive their bill, where the consumption of other residents does not influence its value.

Due to the large number of people living in the city, being able to measure the consumption of each person individually within condominiums can become a challenging activity. For this reason, Sabesp started to make the condominiums themselves responsible for organizing and reading consumption, where they generate monthly reports.

This scenario generates a large demand from condominiums that seek to outsource this work, causing different companies to be created proposing different solutions to be able to carry out such services. Nowadays there are solutions based on smart water meters, which measure and send data to a server by themselves. There are also solutions based on Bluetooth protocols, allowing the meters to store consumption information, and once a person passes by with a cell phone, the data is sent and stored.

However, there are still many scenarios that have not fully adapted to these new technologies, having standard model water meters and thus resulting in the only possible measurement being that of a person performing it. The present work explores this topic having consulted and followed the case of an employee who performs this role, where some problems related to this approach were found.

In real scenarios, the person measuring these water meters chooses to take photographs, where in a month of work there are around 6 thousand photos taken. This scenario generates a very high workload, since the first part of the work involves taking all the photos while the person is present in the condominium, and the second reviewing photo by photo while manually writing down the value of the reading in some database. The second part of the work is taken home or to the office after the day's reading, thus consuming even more hours of the employee's work.

With the advancement of computer vision models nowadays, there are more and more possibilities being explored and analyzed among the most diverse topics, more specifically the field of Optical Character Recognition (OCR), which aims to deal with the extraction and interpretation of text in images or documents. OCR seeks to identify patterns in image pixels and recognize individual characters, then convert them into text that is searchable or editable.

II. Objective

The present work explores OCR applied to the extraction of digits present in water meter photographs, thus managing to obtain the value of the consumption reading, eliminating much of the work of manual annotation. The objective is to analyze and understand the campart of OCR aimed at scenarios similar to the one studied, understanding models and techniques used, as well as proposing new methods and finally training a model to make predictions.

III. Dataset assembly

Present in the photos taken by employees are images taken from blocks, digital panels, blurry or completely black photos, which do not offer any useful information for the problem, and were therefore discarded. Due to the large volume of photos, and the work involved in labeling them, a small sample was chosen in order to validate methods and match the chosen model. Therefore, 200 photos were selected to compose the dataset.

The original images taken by the cell phone are sized 3120x4160 for vertical images, and 4160x3120 for horizontal images. To avoid using a large amount of resources for both storage and training the neural network, the images were resized to 640x640.

By default, images are named based on the dates they were taken, ordering them from most recent to least recent. This way, the data is standardized across condominiums and the types of water meters used. Large sequences of the same type of water meters will be present in the data, which may mean that, when making a cut for training and test data, the model does not have access to all types of water meters in its training. Therefore, to feed these images to the model, the images were randomized and renamed, so that a wide variety of cases are present in the training data, thus increasing the model's generalization capacity.

To compose the training, validation and test data, the photos were distributed in the following order: 120 images for training, 50 for validation and 30 for testing.

In Figure 1 some examples of photos contained within the dataset are shown.

(a) Example 1 (b) Example 2

Figure 1: Dataset photos

IV. Photo analysis

By observing the photos contained in the database, some characteristics of the problem can be noted. Since the water meter works by rotating the digit as water is being consumed, in many cases the digits may be half-displayed. Additionally, lighting conditions, textures and patterns change for each condominium.

It is also possible to notice the repetition of geometric shapes present in the problem. While the hydrometer has a circular shape, the data inside it is arranged in a rectangle shape. This information can be useful for the model, since the vast majority of photos are always rotated.

It is worth mentioning that there are cases where there is no free access to water meters, only the cell phone with the camera can access them, such as ceiling tiles, for example. Therefore, in a possible real application, the solution of limiting the algorithm's operating area to just a small rectangle on the camera, and using a model just to identify digits is not entirely possible, since the user will not be able to aim the cell phone at everyone. the cases.

A. Miscellaneous photos

Along with the most common images of water meters present in the dataset, there are also less-than-ideal cases, with conditions shown in Figure **??**.

(a) Exepmlo 1 (b) Example 2

(c) Example 3 (d) Example 4

Figure 2: Photos

Despite being extreme cases, they demonstrate that noise and changes must be present in the data, both naturally and possibly artificially, meaning that the model can increasingly generalize to cases similar to these.

V. Related works

Since most of the solutions related to computer vision and neural networks have an experimental nature with regard to their architecture, parameters and hyperparameters, a literary research was carried out in order to understand approaches similar to the objective of this work, in order to ultimately explore which models are most used and which other process-escan be applied.

First, a search string S1 was defined, responsible for representing the searched subject. Then, articles were searched among the IEEE and WoS article databases. No filters have been added to searches.

S1: ("computer vision" OR "machine learning" OR "deep learning" OR "neural network" OR "deep neural network" OR "CNN" OR "convolutional neural network" OR "digit recognition") AND ("hydrometer" OR "water meter" OR "water reading" OR "water consumption" OR "water monitoring")

43 scientific articles were returned, of which only 6 were in line with the research proposal, these being:

- Application of Deep Residual Neural Network to Water Meter Reading Recognition [1];
- Research on water meter reading recognition based on deep learning [2];
- Automatic wheel-type water meter digit reading recognition based on deep learning [3];
- Automatic Water Meter Reading Development Based On CNN and LoRaWAN [4];
- Key point localization and recurrent neural network based water meter reading recognition [5];
- Evaluation of Recognition of Water-meter Digits with ApplicationPrograms, APIs, and Machine Learning Algorithms [6].

After selecting and reading the articles, more information about the study area was noted. Table 1 briefly shows how each article addresses the topic, while Table 2 shows general information on the results of each one. It was also possible to understand the following topics.

A. *One-Stage and Two-Stage Models*

There are two types of models implemented for the purpose of this work. The first is One-Stage, which converts the target frame positioning problem into a regression problem, and then performs object classification. In other words, there is no use of pre-generated region proposals (bounding boxes of candidate objects). Some models of this type are YOLO and SSD.

Two-Stages refers to an algorithm that first generates a series of candidate frames as samples by an algorithm and then classifies the samples through a convolutional neural network. Common algorithms include R-CNN, Fast R-CNN, Faster and etc.

The trade-off between the two models is that while One-Stage models are faster, they tend to perform worse than Two-Stage models.

B. *Long Tail Distribution*

Water meters come with a zero reading from the factory, and it takes a long time to consume all their digits from zero, for this reason an expected behavior in the dataset is a Long Tail distribution, in which the digits closest to zero have more instances than the digits close to nine.

Given this, the model will have more resources to learn those classes of digits closer to zero, while the digit nine will have a worse classification performance, for example.

This problem is addressed by Zhu et al.[4] where through the use of Mix Suppression, each digit on the water meter can be changed by exchanging it for other digits of the same image, copying and pasting one digit into another. In other words, the dataset is balanced by replacing the digits, where in a water meter image it could have the digit zero, it is later replaced by the digit six from the same image, for example.

VI. Analyzing the Data

Within the 200 photos labeled by the dataset, the data distribution can be seen in Figure **??**.

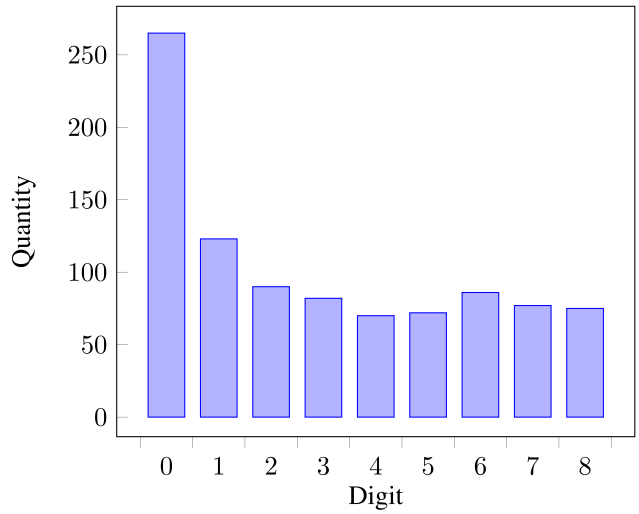

Figure 3: Distribution of digits

Along with the concept of Mix Suppression used by [4], the same concept was now developed and applied to images of rotated water meters. Each digit present in an image has a chance of being selected to be replaced, and once this happens, the remaining digits form a kind of roulette wheel, where a random pointer is rotated, selecting the digit that will be used in the replacement. In other words, the number most present in the dice is more likely to be replaced, and the number less present is more likely to be chosen to replace. Figure **??** shows the use of the developed method.

(a) Example 1

(b) Exexample 2

(c) Example 3

(d) Example 4

Figure 4: Mix Suppression variation developed

After applying this method, the distribution of numbers in the dataset became that of Figure 5.

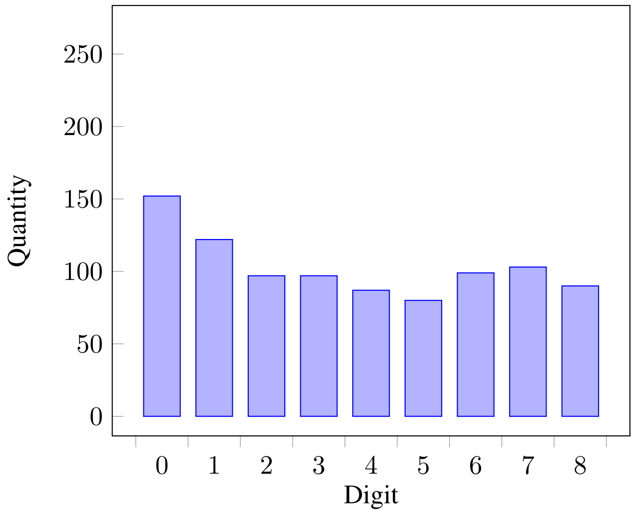

Figure 5: Distribution of balanced digits

It is still possible to notice a certain variation in the quantity of numbers, however, when using the proposed method, as the dataset expands in size, it is increasingly expected that the data will remain more balanced.

VII. Model choice

Regarding the results obtained in table III, it is important to highlight that given the fact that each article works with its own dataset, and there is a wide variation in types of water meters in the world, each with its own dirty conditions and lighting, there is bias present in the results of each article. Those models that perform better may have just one type of water meter, without dirt, illuminated and with digits always visible, thus facilitating the application of a model. Therefore, care must be taken not to relate the best model for one article as the best model for any other.

Among the models found in literary research present in table III, the YOLO model, together with other methods, obtained a consistently satisfactory result. Therefore, since in a possible real-world application the performance and speed of the model are important characteristics, YOLOv8m-OBB was chosen as the implementation model for this article.

The fact that it is the Oriented Bounding Box (OBB) version is due to the fact that the dataset is labeled using this data pattern, that is, rotated rectangles representing each digit and its class.

VIII. Fine tuning

For comparison purposes, YOLOv8m-OBB was trained on both datasets, the first unbalanced, and the second with the Mix Suppression variation applied and balanced data. The practice of using a previously constructed and trained model, using its structure to propose a new problem, is called fine tuning.

IX. Results

The table I shows that small improvements in results were found when the data were balanced by the developed method. The low improvement is possibly coming from the low number of images in the dataset. However, the proposed method proves to be effective and brings satisfactory results, and can be applied to larger data sets generating more impact.

Table I: Results

	P (%)	**R (%)**	**mAP50 (%)**	**mAP50-95 (%)**
Unbalanced	0.81315	0.73608	0.81599	0.67963
Balanced	0.83994	0.7735	0.84313	0.70714

X. Conclusion

In this work, it was possible to understand a new area of computer vision: the application of OCR to water meters. Features present in the nature of the problem were discovered and analyzed. It was also possible to carry out literary research with the aim of exploring methods and models used by the academic community, as well as proposing a variation of an existing method, resulting in significant improvements to the chosen YOLOv8m-OBB model. This study contributes to the advancement and improvement of the application of OCR in water meters, offering insights for future research and developments in this specific area of computer vision.

Table II: Summary observations of the articles

Article	Note
Application of Deep Residual Neural Network to Water Meter Reading Recognition	First use ResNet-101 for feature extraction; then uses the RPN network to generate ROI regions; then divides the ROI into nine regions, generates the feature spectrum of each region, and finally performs voting classification detection.
Research on water meter reading recognition based on deep learning	There are two plans proposed, plan A uses a sequence of steps that includes adjusting the size of the input image, identifying the numbers on the water meters and outputting the numbers after sorting. On the other hand, Plan B involves identifying the black rectangle, cropping the images as input, and outputting the numbers after sorting.
Automatic wheel-type water meter digit reading recognition based on deep learning	Mix suppression was used to expand the dataset, improve class distribution, and increase small object detection performance. Attention mode BSAM was implemented together to guide the CNN's attention to the object concentration area and further improve the performance of the object detection model.
Automatic Water Meter Reading Development Based On CNN and LoRaWAN	There is no explicit mention of the performance metrics or accuracy of the models. The article focuses on IoT technology, LoRaWAN networks, convolutional neural networks (CNN) and their application in developing an intelligent system for automatically reading analog water meters.
Key point localization and recurrent neural network based water meter reading recognition	Key point location and water meter reading recognition method based on recurrent neural network. three main modules, including locating key points in the digital area, correcting and segmenting distorted images, recognizing and transcribing the water meter reading sequence.
Evaluation of Recognition of Water-meter Digits with Application Programs, APIs, and Machine Learning Algorithms	Considering the context of digit detection in water meters, compare correction of 5 programs/APIs, Anyline, Line, Google Vision, Microsoft Azure Computer Vision, and Naver (Clovar) with solutions such as KNN, SVM, CNNs, AutoML , Tesseract and etc. The paper showed that the CNN-based architecture can provide recognition results comparable to commercial programs and APIs. Photos classified into three classes: straight, 90 degrees to the right, upside down

Table III: Literature search results

Article	Own Dataset	Dataset size	Box extraction methods used	Used target-detection architectures	Evaluation of model(s)	
					Acc (%)	mAP (%)
Application of Deep Residual Neural Network to Water Meter Reading Recognition	Yes	160,785 images (15,000 training and 145,785 test)	RPN	R-FCN	81.79	
Research on water meter reading recognition based on deep learning	Yes	4000 original images 14,400 with expansion (12,960 training and 1,440 test)	RPN	SSD, YOLOv3 and Faster R-CNN	Plan A SSD: 78.63 YOLOv3: 87.30 Faster R-CNN: 82.96 Plan B SSD: 80.42 YOLOv3: 90.61 Faster R-CNN: 81.53	Plan A SSD: 75.67 YOLOv3: 78.10 Faster R-CNN: 77.07 Plan B SSD: 78.61 YOLOv3: 82.02 Faster R-CNN: 75.46
Automatic wheel-type water meter digit reading recognition based on deep learning	Yes	1277 images (700 training and 577 test)		Yolov4-tiny (with Mix suppression and BSAM)		97.2
Automatic Water Meter Reading Development Based On CNN and LoRaWAN				FCSRN		
Key point localization and recurrent neural network based water meter reading recognition	Yes	45861 (43,681 training and 2,000 tests)	Stacked hourglass networks	CRNN (with distortion correction)	95.3	
Evaluation of Recognition of Water-meter Digits with Application Programs, APIs, and Machine Learning Algorithms	Own dataset, MNIST, EMNIST and SVHN	96 original images, mixed in the future with other datasets		Anyline, Line, Google Vision, Microsoft Azure Computer Vision, Naver (Clovar), AutoML, Tesseract, KNN, SVM eCNNs	(Top 3 models from each photo class) Straight Anyline: 96.88 Naver OCR: 92.19 GoogLeNet:86.72 90 degrees to the right Naver OCR:87.5 GoogLeNet:78.91 Automl: 75 Head end Naver OCR: 85.16 GoogLeNet: 74.22 Microsoft Azure: 69.53	

Empty cells represent unreported data.

REFERENCES

[1] Y. Peng and Z. Chen, "Application of Deep Residual Neural Network to Water Meter Reading Recognition," 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2020, pp. 774-777, doi: 10.1109/ICAICA50127.2020.9182460.

[2] Liang, Y., Liao, Y., Li, S. et al. Research on water meter reading recognition based on deep learning. Sci Rep 12, 12861 (2022). https://doi.org/10.1038/s41598-022-17255-3

[3] Zhu, J., et al. "Automatic wheel-type water meter digit reading recognition based on deep learning," in Journal of Electronic Imaging, vol. 31, pp. 023023, 2022.

[4] Bangkit, H., et al, "Automatic Water Meter Reading Development Based On CNN and LoRaWAN," in Book title is required!, 2023, pp. 212-215.

[5] Jiguang Zhang, undefined., et al. "Key point localization and recurrent neural network based water meter reading recognition," in Displays, vol. 74, pp. 102222, 2022.

[6] Eurviriyanukul, K., et al, "Evaluation of Recognition of Water-meter Digits with Application Programs, APIs, and Machine Learning Algorithms," in 2020 8th International Electrical Engineering Congress (iEECON), 2020, pp. 1-4.