



Image-based Automatic Dial Meter Reading in Unconstrained Scenarios

Gabriel Salomon^{*}, Rayson Laroca, David Menotti

Department of Informatics, Federal University of Paraná (UFPR), Curitiba, PR, Brazil

ARTICLE INFO

Keywords:

Automatic Meter Reading
Dial meters
Pointer-type meters
Deep learning
Public dataset

ABSTRACT

The replacement of analog meters with smart meters is costly, laborious, and far from complete in developing countries. The Energy Company of Paraná (Copel) (Brazil) performs more than 4 million meter readings (almost entirely of non-smart devices) per month, and we estimate that 850 thousand of them are from dial meters. Therefore, an image-based automatic reading system can reduce human errors, create a proof of reading, and enable the customers to perform the reading themselves through a mobile application. We propose novel approaches for Automatic Dial Meter Reading (ADMR) and introduce a new dataset for ADMR in unconstrained scenarios, called UFPR-ADMR-v2. Our best-performing method combines YOLOv4 with a novel regression approach (AngReg), and explores several post-processing techniques. Compared to previous works, it decreased the Mean Absolute Error (MAE) from 1343 to 129 and achieved a meter recognition rate (MRR) of 98.9% — with an error tolerance of 1 Kilowatt-hour (kWh).

1. Introduction

Meter reading is a laborious task and usually requires an employee to visit each meter periodically in order to measure the gas, water and energy consumption of consumers around the world [1–3]. Although the world currently has the technology to perform readings using devices such as smart meters [4], there are still millions of old (analog and electronic) meter devices in operation, even in developed countries [5].

1.1. Research motivation

Replacing all of the deployed meters is an expensive and time-consuming procedure, still far from complete [5,6].

There are two main categories of energy meters: (i) analog (i.e., with cyclometer and dial displays) and (ii) digital (i.e., with electronic display, and smart meters) [5]. Most works in the literature focus on digit-based meters (cyclometer and digital ones) [7–9]. However, the Energy Company of Paraná (Copel) (presented in Section 3) performs more than 4 million energy meter readings per month, of which we estimate that 850 thousand are from dial meters (based on statistical sampling of 1 million images provided by Copel). This means that there is still a significant percentage of these meters in use. Regarding the automatic reading of pointer-based meters (dials and gauges), most works focus on industrial gauges such as pressure meters, ammeters, and voltmeters [10–13].

Automatic Dial Meter Reading (ADMR) stands for the automatic reading of dial-based (or pointer-based) meters.

In this work, we tackle the issue of reading multi-dial meters using images (see Fig. 1) since it requires no change to the meters, being a faster, easier and cheaper solution to deploy than smart meters, which require the deployment of a data network (wired or wireless) and the replacement of the meter.

1.2. Literature review

Although there are many published works focused on image-based ADMR systems, most of the proposed methods are designed and evaluated exclusively in well-controlled environments [10,14,15], i.e., on images with no challenging factors that are found on a daily basis by readers such as inaccessible meters (leading to oblique views), variable lighting conditions, distant meters, reflections, occlusion, blur, and wear caused by weather, rust and dust. This is in stark contrast to related research areas, where in recent years the research focus shifted to unconstrained scenarios [7]. In addition, the majority of works do not provide a public dataset in order to enable reproducibility and fair comparison between approaches [10,14,15]. There is also a lack of state-of-the-art methods being explored, with a lot of handcrafted methods being used [16–19]. This is one of the rare works to employ deep learning models for the entire ADMR pipeline.

1.3. Necessity of the research

Considering electricity losses [20], according to the 2019 annual report of the Brazilian Electricity Regulatory Agency (ANEEL) [21],

^{*} Corresponding author.

E-mail addresses: gspaniceto@inf.ufpr.br (G. Salomon), rblsantos@inf.ufpr.br (R. Laroca), menotti@inf.ufpr.br (D. Menotti).



Fig. 1. Representative samples from the proposed UFPR-ADMR-v2 dataset. All images were taken in unconstrained scenarios, with significant variations in lighting, distance, contrast, noise, rotation and camera.

approximately 46% of the company's non-technical electricity losses are caused by issues such as theft, tampering with meters, non-working meters, reading and billing errors. It is unclear; however, the occurrence ratio for each of the aforementioned events. In absolute values, non-technical losses represent 33.3 TWh per year (6.6% of the generated electricity).

In this regard, a robust image-based Automatic Meter Reading (AMR) system can speed up the reading process, reduce human errors, create proof of reading, and enable the customers to perform the reading themselves through a mobile application, eliminating the need for employees of the service company traveling around to perform local meter reading at each consumer unit.

1.4. Novelty and main contributions

In summary, the main contributions of this work are:

- A public dataset for ADMR, called UFPR-ADMR-v2 (see Fig. 1), that includes 5000 fully-annotated images of dial meters (totaling 22,410 annotated dials) acquired in unconstrained scenarios by the service company's employees themselves.¹ The proposed dataset contains a well-defined evaluation protocol to assist the development of new approaches toward this problem as well as to enable a quantitative and qualitative comparison among published works. To the best of our knowledge, this is the largest publicly available dataset in the ADMR literature;
- A novel approach toward image-based ADMR that combines a detection model with our proposed regression approach (AngReg) in order to enable error correction for the most significant dials, thus decreasing the Mean Absolute Error (MAE) considerably. We remark that most of the reading errors made by this approach occurred in challenging cases, where even humans can make mistakes due to artifacts in the counter region;

¹ The UFPR-ADMR-v2 dataset is publicly available to the research community at <https://github.com/guesalomon/ufpr-admr-v2-dataset/>. Access is granted upon request, i.e., interested parties must register by filling out a registration form and agreeing to the dataset's terms of use.

- The most extensive comparative assessment in the ADMR literature, with 20 approaches (18 models) being adjusted and evaluated under exactly the same conditions. We are not aware of any work in the literature where so many methods were evaluated in the experiments.

A preliminary version of this work was published at the 2020 International Joint Conference on Neural Networks (IJCNN) [5]. This work differs from that in several aspects. Here, we present and assess novel approaches for ADMR, such as regression and segmentation-free methods. We also evaluate more detection-based methods, such as YOLOv4 [22] and the end-to-end approach proposed in [7]. In addition to expanding the UFPR-ADMR-v1 dataset considerably, with new images (from 2000 to 5000) and refined annotations, we provide solutions for each of the primary causes of reading errors found in [5]. In this way, we attain much better results than those reported in our previous work, especially considering that 80% of the reading errors made by the best performing approach – YOLOv4 (R) + (X) – occurred on the least significant dial. In fact, 85% of all errors have MAE = 1. With an error tolerance of 1 Kilowatt-hour (kWh), our final system achieves an impressive meter recognition rate (MRR) of 98.9%.

1.5. Organization of the paper

The remainder of this paper is organized as follows. We review related works in Section 2. The UFPR-ADMR-v2 dataset is introduced in Section 3. The proposed approaches are presented in Section 4. In Section 5, we report the experiments performed and discuss the results obtained. Finally, conclusions and future work are given in Section 6.

2. Related work

Most works in the AMR literature focus on digit-based meters (cyclometers and digital meters). The literature for ADMR is more scarce. Most of the works on pointer recognition focus on industrial gauges, with handcrafted methods and well-controlled environments regarding lighting, weather, and image quality; some of them used only images from the same distance and angle [12,16,18,23]. Recently, deep learning approaches have gained traction in the AMR [7,24–26] and ADMR [5,10,13,15,27] contexts. In this section, we review some relevant works on ADMR. We also briefly describe some digit-based approaches that achieved promising results. Lastly, we conclude this section with final remarks.

Several works explored handcrafted features, such as Hough Transform (HT) [16–18,28], to locate the dials or the pointers. The steps usually include image binarization on the preprocessing stage, Hough Circle Transform (HCT) for dial detection, and pointer angle detection using Hough Line Transform (HLT) or similar methods. These approaches generally work well in constrained environments but may not be suitable for real-world scenarios (i.e., with uneven lighting and the presence of noise) [7,29].

Mask R-CNN was proposed for reading meters based on pointers in [10,15,30]. Fang et al. [15] used it to find reference key points on both the pointer and the gauge scale marks, whereas He et al. [10] and Zuo et al. [30] focused on segmenting the dial and the pointer. In these three works, the angle between the pointer and the respective dial was explored to retrieve the meter reading. Although promising results were achieved in all of them, only private datasets were used in the experiments.

Hou et al. [27] explored the use of a lightweight CNN model to read cropped industrial gauges (the detection of the region of interest was not addressed in their work). On the other hand, region-based Fully Connected Networks (FCNs) were explored for meter detection by Huang et al. [12], which applied handcrafted methods in the remainder of the pipeline.

Liu et al. [11] explored four Convolutional Neural Network (CNN)-based object detectors (i.e., Fast R-CNN, Faster R-CNN, Single Shot

MultiBox Detector (SSD), and YOLO) for meter detection. High mean Average Precision (mAP) values were achieved for this task, however, no deep learning-based approach was evaluated at the recognition stage.

In [31], Sowah et al. extracted high-level features using a cascade of contour filters and performed the final recognition using a regression model and K-Nearest Neighbors (KNN). Although only handcrafted features were explored, it is noteworthy that this is one of the few works where experiments were carried out with different models of gauges.

Cai et al. [13] and Howells et al. [14] tackled the lack of publicly available gauge datasets by introducing approaches to generate synthetic data. Cai et al. [13] first generated synthetic samples by varying the center and angle of the pointer and then trained a CNN model to perform regression of the pointed values. Howells et al. [14] leveraged synthetic 3D-rendered gauge images to train two CNN detectors, one for gauge detection and one for gauge reading. Their system corrects each gauge's perspective and identifies the scale marks to predict the reading correctly. Although their system outperformed other methods for gauge reading, it was evaluated exclusively on images with the gauge well centered and occupying a large portion of the image, i.e., the images used in their experiments are not as unconstrained as those collected in the field by the service company's employees.

The following approaches focused on multi-dial meters, more specifically energy meters. Vega et al. [19] exploited handcrafted features to perform dial recognition in energy meters. Methods such as binarization and line intersection were used to extract features and the Scale-Invariant Feature Transform (SIFT) was employed to detect the counter region. This method was evaluated on a private dataset containing only 141 well-controlled images.

In [23], Tang et al. proposed a complete framework for dial energy meter reading based on binarization, line intersection, and morphological operations. As in [19], the images used in the experiments were obtained in a well-controlled environment and belong to a private dataset.

In our previous work [5], we compared several deep learning-based methods for ADMR. More specifically, we fine-tuned object detectors (e.g., Faster R-CNN [32] and YOLOv3 [33]) for dial recognition directly on the input images, i.e., without the need for Region of Interest (ROI) detection or segmentation. Based on our experimental evaluation, performed on images from the UFPR-ADMR-v1 dataset, we highlighted the trade-off between precision and speed achieved by YOLOv3 since it reached recognition rates competitive to Faster R-CNN while being able to process three times faster.

There are some digit-based AMR research worth mentioning [7, 24, 34]. Gómez et al. [34] introduced a CNN-based segmentation-free approach to perform digit-based meter reading, without the need for counter region detection. They achieved promising results in a private dataset. Nonetheless, the authors only evaluated their method against traditional algorithms that rely on handcrafted features, which are not robust to unconstrained scenarios [7]. Laroca et al. [24] developed a two-stage approach to tackle AMR. Fast-YOLOv2 [35] was employed for counter detection and three distinct CNNs were evaluated in the recognition stage. As their two-stage approach failed in some cases where the counter region was tilted, in [7] authors from the same research group designed a multi-task network to detect the four corners of the counter region to rectify it before the recognition stage. In their experiments, the improved system outperformed all baselines in terms of recognition rate while still being quite efficient.

2.1. Remarks

Most of the approaches were designed specifically for reading single-dial meters, which are also known as gauges. Even though the problems are similar, there are three fundamental differences between them: (i) a small error in a multi-dial meter may result in a substantial reading error (in case the error(s) occur in the most significant dials);

(ii) there are intersections between dials in several meters, which may interfere with the reading process and segmentation-based methods; (iii) in gauges, the rightmost dials cannot be used for correction (as gauges usually only have one dial). More details on this type of correction are provided in Section 5.

Most of the works performed experiments exclusively on private datasets. Although there are a few publicly available datasets for AMR (e.g., [7, 8, 24, 26, 36, 37]), none of them have images containing pointer-type meters, only digit-based ones. As far as we know, the only dataset with images of multi-dial meters currently available in the literature is UFPR-ADMR-v1, which we introduced in the preliminary version of this work [5]. Besides, the majority of works deal with well-controlled and constrained scenarios, which are immensely simpler than real-world scenarios.

As mentioned in Section 1, there are several differences between this work and [5]. For example, this work presents and assesses novel approaches for ADMR, such as regression and segmentation-free methods. We emphasize that, to the best of our knowledge, no previous pointer-based works explored Long Short-Term Memory (LSTM)-based approaches to tackle the reading process as a sequence problem. We consider this a promising approach that can leverage the information of the previous dial to predict the current one, instead of reading each dial as an independent instance.

We also evaluate more detection-based methods, such as YOLOv4 [22] and the end-to-end approach proposed in [7], and expand the UFPR-ADMR-v1 dataset considerably (from 2000 to 5000 images). Lastly, we tackle the primary sources of errors found in [5], providing solutions to them (i.e., we explored data augmentation and post-processing techniques to reduce errors), thus increasing the MRR from 75.25% to 92.50% and reducing the MAE from 1343.29 to 187.

We reinforce that most works in the ADMR literature do not provide error analyses. There are also not many post-processing methods available in the literature, that is, some works perform angle corrections, but none of them proposed a more robust multi-dial correction method.

3. The UFPR-ADMR-v2 dataset

In our previous work [5], we introduced the UFPR-ADMR-v1 dataset, which contains 2000 images acquired in the field by the employees of the Energy Company of Paraná (Copel), a company of the Brazilian electricity sector that serves more than 4 million consuming units per month [24, 38].

In this work, we introduce an expanded version of that dataset, called UFPR-ADMR-v2, with 3000 more images – totaling 5000 images (with 22,410 dials) – and new annotations such as the approximated pointed value labels (with one decimal place precision) and segmentation masks for the dials. To the best of our knowledge, this is the largest and most complete publicly available dataset in the ADMR literature. Some representative samples are shown in Fig. 1 (in Section 1).

All 5000 images were obtained on-site by Copel employees as proof of reading. For this task, cell phones of various brands were used, most of them with low-end cameras. The images were compressed and resized for storage, thus reducing their overall quality. In this sense, the images have a resolution of 640×480 or 480×640 pixels, with an average size of 60 kB.

There are many challenges a robust system must be able to handle in order to correctly perform meter readings in real-world scenarios. As illustrated in Fig. 2, these challenges include inaccessible meters (resulting in oblique views), variable lighting conditions, distant meters, reflections, occlusion, blur, and the wear caused by weather, rust and dust. Furthermore, dial meters are subject to mechanical failure such as misalignment, non-working pointers, among others, which undermine the performance of an ADMR system.

The labeling process was a time-consuming task, lasting more than 350 h in a custom-made tool — not including the time required to improve the tool itself as well as the relabeling of several images.

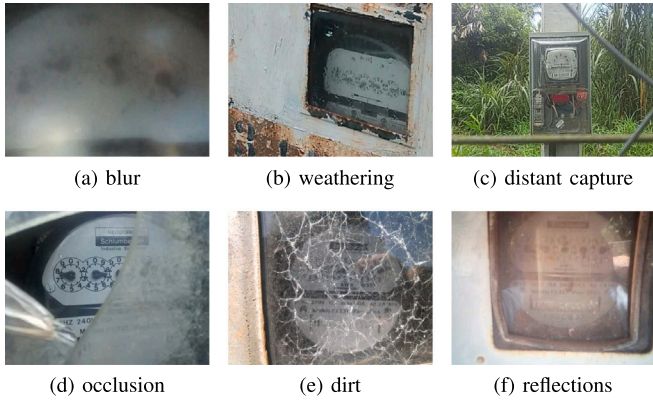


Fig. 2. Some of the challenging scenarios in the UFPR-ADMR-v2 dataset. For evaluation purposes, we only selected images where it is possible for a human to recognize the correct meter reading.

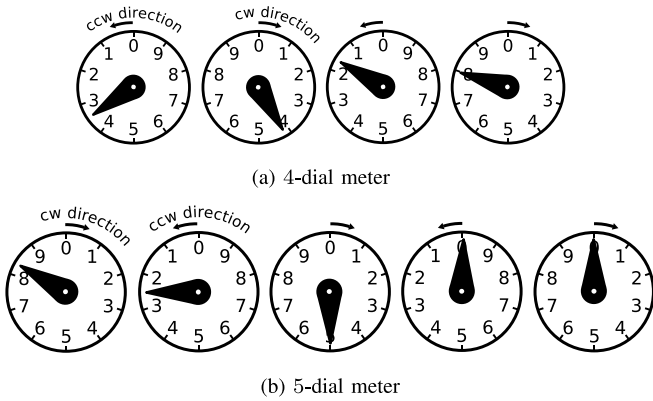


Fig. 3. Illustration of the energy meter dials present in the proposed dataset. Observe the alternating direction on sequential dials, with the least significant dial (i.e., the rightmost one) always oriented clockwise.

During the labeling process, we identified a pattern regarding the direction of the dials (clockwise (CW) or counterclockwise (CCW)) in all meters in the proposed dataset: the dials have alternating directions, with the least significant dial – i.e., the rightmost one – always oriented clockwise. Thus, the orientation is always (CW, CCW, CW, CCW, CW) for the 5-dial meters and (CCW, CW, CCW, CW) for the 4-dial meters, as illustrated in Fig. 3. This information can be useful for training methods focused on the angle of the pointer, regardless of its orientation. In other words, it is possible to correctly read the dials even in cases where the digits are not legible, that is, considering only the angle of the pointer, since the orientation can then be retrieved using the number of detected dials and their index on the counter region.

Fig. 4 shows the distribution of the dial classes in the dataset, with one decimal place precision. Note that there is an almost uniform distribution across the digits; however, the values pointed most often are close to the digits themselves and far from the middle position between two digits.

3.1. Evaluation protocol and metrics

Following the protocol defined in [5], the proposed dataset is divided into three disjoint subsets: 3000 images for training (60%), 1000 images for validation (20%), and the remaining 1000 images for testing (20%). The subsets are explicitly defined along with the UFPR-ADMR-v2 dataset, as it is common in the AMR literature [5,7,26]. Note that we maintained the original split for the images from the first version of the dataset, i.e., the UFPR-ADMR-v1's training images

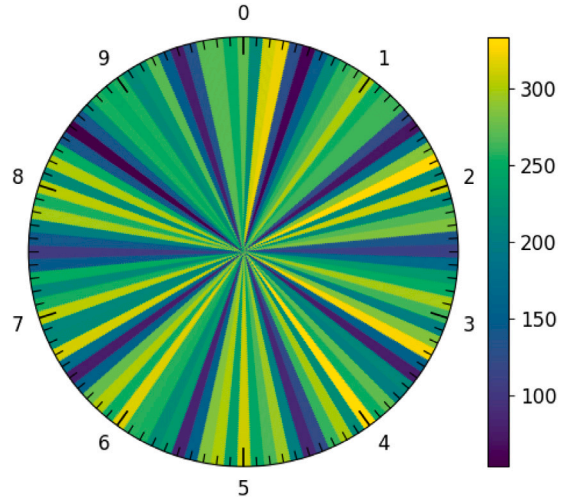


Fig. 4. Distribution of the pointer positions in the UFPR-ADMR-v2 dataset.

are in UFPR-ADMR-v2's training subset, and the same is true for the validation and test subsets.

We propose the use of three metrics for evaluation: (i) meter recognition rate (MRR), (ii) dial recognition rate (DRR), and (iii) Mean Absolute Error (MAE). In the following paragraphs, we detail them.

The MRR is obtained by comparing, for each of the N meters, the predicted sequence ($pred_m$) with the ground-truth sequence (gt_m):

$$MRR = \frac{1}{N} \sum_{m=1}^N match(pred_m, gt_m) \quad (1)$$

$$match(x, y) = \begin{cases} 1, & \text{if } x = y, \\ 0, & \text{if } x \neq y. \end{cases}$$

The DRR is based on the Levenshtein distance (also known as edit distance), a common method for measuring the distance between two sequences of characters. This metric has the advantage of dealing with certain minor errors (e.g., an undetected dial, resulting in strings of different lengths) without heavily punishing them, as the distance between the predicted sequence and the ground truth is calculated and divided by the longest sequence. Finally, subtracting the error from 1 results in the dial recognition rate for each meter. The DRR corresponds to the mean of the recognition rates:

$$DRR = \frac{1}{N} \sum_{m=1}^N \left(1 - \frac{lev(pred_m, gt_m)(|pred_m|, |gt_m|)}{\max(|pred_m|, |gt_m|)} \right) \quad (2)$$

MAE is useful for penalizing errors in the most significant dials (i.e., the leftmost ones). Considering the sequences as integers, we can obtain the mean absolute error as follows:

$$MAE = \frac{1}{N} \sum_{m=1}^N |p_m - g_m| \quad (3)$$

4. Evaluated approaches

In this paper, we evaluate three different approaches for ADMR: (i) object detection with recognition; (ii) object detection with regression; and (iii) segmentation-free recognition. The pipeline for each approach is illustrated in Fig. 5.

4.1. Object detection with recognition

As can be seen in Fig. 5(a), this approach consists of using object detection networks to simultaneously detect and classify the dials directly in the input image. The networks are trained to predict 10 classes

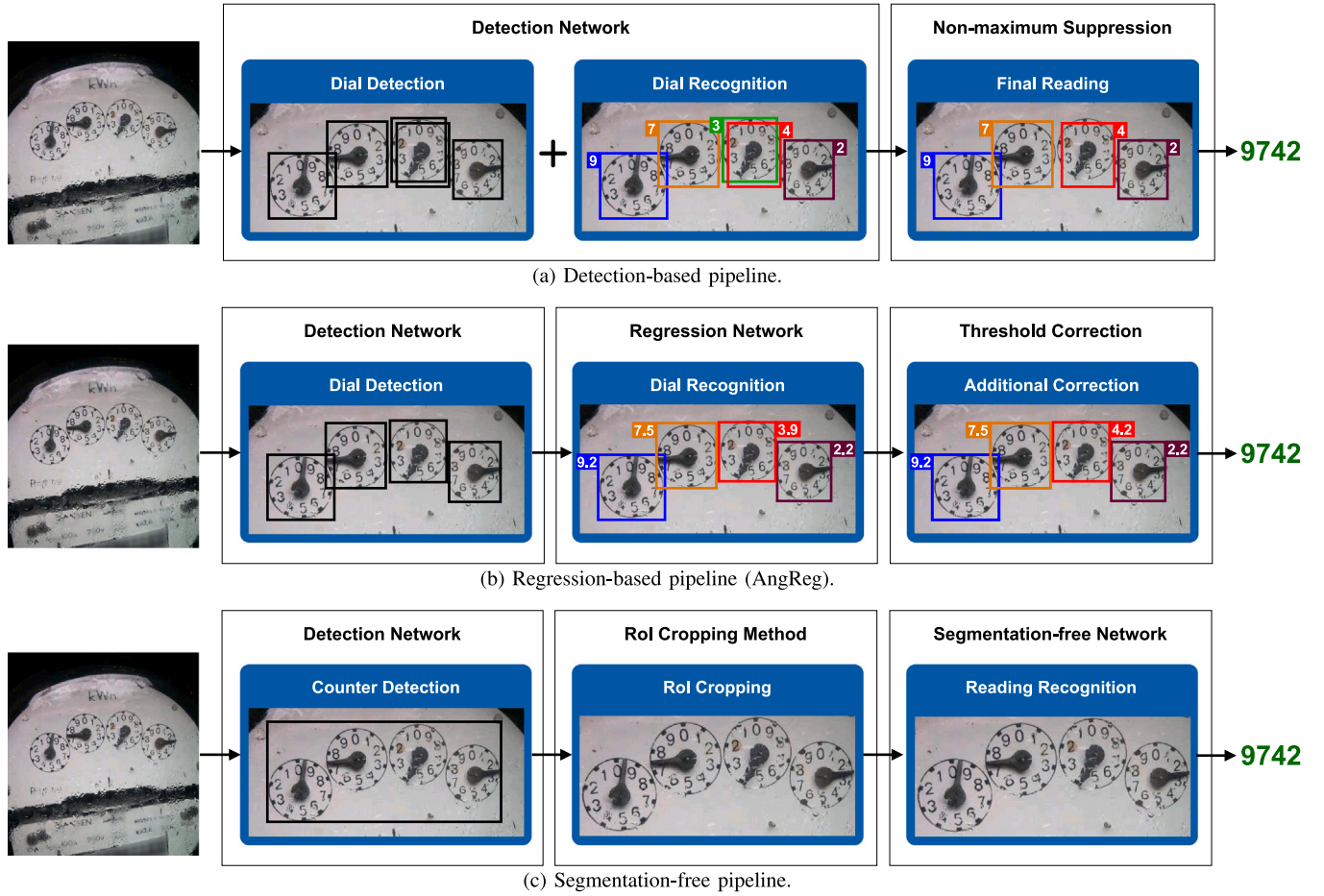


Fig. 5. The proposed pipelines for ADMR. Note that in the detection network the images are zoomed-in for better viewing. In fact, the dial and counter detection are performed on the entire image.

(i.e., 0–9) using the bounding box of each dial as input. Afterward, we apply a Non-Maximum Suppression (NMS) algorithm to eliminate redundant detections (those with Intersection over Union (IoU) ≥ 0.5) since the same dial might be detected/recognized more than once by the network.

For this approach, we decided to evaluate YOLOv3 [33] and YOLOv4 [22] considering that YOLO-based models been successfully employed for various computer vision tasks [39,40] and have achieved promising reading results in images of digit-based meters [7,41].

Shallower detection models, such as those used by Laroca et al. in [24] (CR-NET) and in [7] (Fast-OCR), also fit this approach. However, given the smaller size of these networks, they are fed with the counter region previously detected by another detection network and not directly in the input image.

4.2. Object detection with regression (AngReg)

We introduce a novel regression approach, called AngReg, for ADMR. AngReg comprises a detection model, followed by a custom regression model (with trigonometric activation functions) and a novel post-processing correction approach.

Similar to the object detection approach, AngReg also leverages an object detector (i.e., YOLOv4 [22]) to obtain the position of all dials from the input image (see Fig. 5(b)).

After the detection stage, fine-tuned ResNet [42] and Xception [43] models are explored to extract image features, and a custom FCN with the hyperbolic tangent functions is used to predict the reading for each detected dial. The AngReg topology is described in more detail in Appendix A.

ADMR can be addressed as a regression problem since we can introduce more precision to the pointer prediction (e.g., 5.75, while object detection approaches usually yield a discrete label (e.g., ‘5’). This added precision enables the application of our correction method (described below), which provides information on whether the next dial’s prediction matches the current one based on the decimal place (i.e., if the decimal place is too far from the integer part of the next dial, probably one of the predictions is not accurate).

For training the AngReg regression models, we evaluated several target functions such as linear, sigmoid, and hyperbolic tangent. The labels were converted according to the activation function for the sigmoid and tangent targets. For the sigmoid function, we scaled the label values between 0.0 and 1.0. We converted the labels values to angles for the tangent function and used the sine and cosine to obtain the polar coordinates correctly. The encoding of the angles is performed using the following formulation to obtain Cartesian coordinates from polar coordinates (assuming the radius is 1):

$$\theta \mapsto (\sin(\theta), \cos(\theta)). \quad (4)$$

After this conversion, the target function is composed of two hyperbolic tangent perceptrons (one for the sine coordinate and the other for the cosine coordinate — the y and x coordinates, respectively). To predict the correct angle for a cropped dial, we need to convert those two Cartesian coordinates to polar coordinates (in our case, to an angle θ).

To convert the coordinates back from Cartesian to polar we use the following formulation:

$$(y, x) \mapsto \arctan 2(y, x) \quad (5)$$

$$\arctan 2(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{if } x > 0, \\ \frac{\pi}{2} - \arctan\left(\frac{x}{y}\right) & \text{if } y > 0, \\ -\frac{\pi}{2} - \arctan\left(\frac{x}{y}\right) & \text{if } y < 0, \\ \arctan\left(\frac{y}{x}\right) \pm \pi & \text{if } x < 0, \\ 0 & \text{if } x = 0 \text{ and } y = 0. \end{cases}$$

After the predicted angle is yield, we can obtain the numerical value for the prediction (ranging between 0.0 and 9.9) using the orientation pattern (see Section 3) to determine whether the dial is clockwise or counterclockwise oriented.

After the numerical value is retrieved, we apply a heuristic method to correct certain reading errors. This method uses the dial immediately on the right to check whether the predicted value for a given dial is accurate. For instance, if the current dial's prediction is 3.9, but the next dial's prediction is 2.2, it is very likely that the correct reading for the current dial is 4 instead of 3. This case is illustrated in Fig. 5(b). Note that the rightmost dial (the least significant one) cannot be corrected using this method, as there is no dial to the right to imply the correct value.

We used the validation subset to find the best thresholds for this correction method. These thresholds are basically two pairs of values (i.e., the lower and upper bounds for a given dial and the one immediately to the right) that indicate whether to correct the prediction for the current dial using the reading value predicted for the dial to the right.

Fig. 5(b) shows an example where this correction method is applied. Observe that the dial pointing to 3.9 is followed by a dial pointing to 2.2 (these values were predicted by the regression model). Suppose the upper correction threshold is $N.75$ for the dial being analyzed, and the lower correction threshold for the next dial is $N.5$. As $3.9 > N.75$ (considering only the decimal values) and $2.2 < 2.5$, we correct the predicted value for the dial being analyzed to 4.2. Lastly, the float values are rounded to integers to produce the final reading.

4.3. Segmentation-free recognition

The segmentation-free approach refers to the meter reading being predicted holistically from the counter region, i.e., deep models are employed to predict the meter reading without the need for detecting/segmenting each of its dials. Such models have been successfully explored in Optical Character Recognition (OCR) applications [44–46].

For this approach, it is necessary to first detect the ROI (see Fig. 5(c)) due to the fact that in unconstrained scenarios the ROI can occupy a very small portion of the input image — this stage is commonly referred to as “counter detection” in the AMR literature. For this task, we used a lightweight detection network [7], called Fast-YOLOv4-SmallObj, as it was robust enough to reach 100% F-score (IoU ≥ 0.5) in our experiments.

We then evaluated several segmentation-free models well known in the literature. Some of them are multi-task networks, i.e., those proposed or adapted by Gómez et al. [34], Laroca et al. [24] (Multi-task), and Calefati et al. [44], while others are recurrent-, sequential- or Transformer-based networks commonly employed for scene text recognition: CRNN [47], GRCNN [48], R²AM [49], RARE [50], Rosetta [51], STAR-Net [52], TRBA [53], and ViTSTR-Base [45].

4.4. Data augmentation

We explored various data augmentation strategies to train the evaluated networks, such as random cropping, conversion to grayscale, and random perturbations of hue, saturation and brightness. Our main goal with data augmentation is to reduce lighting and symmetry errors (those errors occur due to the lack of sufficient and heterogeneous data). Thus, we chose the most common lighting, color and position

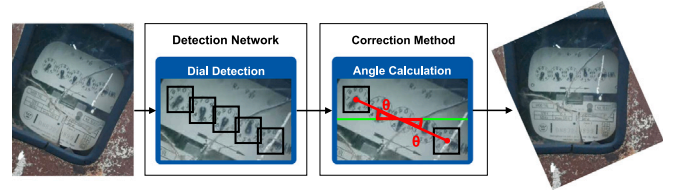


Fig. 6. Illustration of the angle correction method. The angle (θ) between the line (in red) that passes through the middle of the first and last detected dials and the x -axis (in green) is computed. Afterward, the correction method rotates the image $-\theta$ degrees in order to make the dials parallel to the x -axis.

data augmentation techniques used in the literature [24,54]. While there are several data augmentation libraries available, we created the modified images using Albumentations [55], which is a well-known Python library for image augmentation that has been successfully used in several applications, including AMR [7].

In preliminary experiments, we also evaluated the possibility of ignoring the dial direction, forcing the methods to learn the pointer angle rather than the digit and dial direction. For this purpose, we changed the labels on the CCW dials to match the CW dials (1 became 8, 2 became 7, and so on). The label can be converted back after prediction using the dial direction pattern described in Section 3. However, contrary to what we expected, there was no significant improvement in the results achieved (further investigation is needed in this regard).

4.5. Post-correction

Considering that the images in the UFPR-ADMR-v2 dataset were acquired in unconstrained environments and, therefore, some images have the meter rotated at an angle of almost 45° , we leverage a post-processing strategy to improve the results obtained by the detection-based approaches. First, for the predictions obtained for each meter, we calculate the angle of the line segment that passes through the middle of the first and last dial bounding boxes. Then, if this angle is greater than a defined threshold (in this work, $\theta > 2.5^\circ$), we rotate the image accordingly so that the angle with the x -axis is approximately 0° ; note that this threshold ($\theta > 2.5^\circ$) was defined based on the results achieved in the validation set. In these experiments, rotating the images in which the angle was less than 2.5° did not improve the results. Lastly, we feed the corrected image to the network and consider these predictions the final reading. An illustration of the whole process is provided in Fig. 6.

5. Experiments

In this section, we report the experiments performed to verify the effectiveness of the evaluated approaches.

5.1. Setup, deep models, and evaluation protocol

All our experiments were carried out on an AMD Ryzen Threadripper 1920X 3.5 GHz CPU, 96 GB of RAM, and an NVIDIA Quadro RTX 8000 GPU. For a fair comparison, we trained all approaches (detection, regression, and segmentation-free) using exactly the same images.

Table 1 presents an overview of the 18 deep models we trained and tested for this work (these models were mentioned throughout Section 4), listing the framework we used to implement each of them. These models were chosen mainly because they have been explored for several tasks with promising/impressive results [34,44,45] and due to the fact that we believe we have the necessary knowledge to train/adjust them in the best possible way in order to ensure fairness in our experiments, as the respective authors provided enough details about the architectures used and because we designed/used similar networks in previous works [5,7,56].

Table 1

Deep models explored in our experiments along with the respective frameworks used to implement them.

Model	Framework
CR-NET [57]	Darknet
Laroca et al. [7]	Darknet
YOLOv3 [33]	Darknet
YOLOv4 [22]	Darknet
Calefati et al. [44]	Keras
Gómez et al. [34]	Keras
Multi-task [56]	Keras
ResNet50 [42]	Keras
ResNet101 [42]	Keras
Xception [43]	Keras
CRNN [47]	PyTorch
GRCNN [48]	PyTorch
R ² AM [49]	PyTorch
RARE [50]	PyTorch
Rosetta [51]	PyTorch
STAR-Net [52]	PyTorch
TRBA [53]	PyTorch
ViTSTR-Base [45]	PyTorch

Each model was trained and tested using either the framework where it was originally implemented or well-known public repositories. In summary, the YOLO-based models (detection networks) were trained using the Darknet framework,² the multi-task models were trained using Keras,³ and the other models were trained using a fork of the open source repository of Clova AI Research,⁴ which was employed to record the 1st place of ICDAR2013 focused scene text and ICDAR2019 ArT, and 3rd place of ICDAR2017COCO-Text and ICDAR2019 ReCTS (task1). We provide more details on the implementation of these methods in Appendix B.

For the detection networks, we used the *mAP* evaluation metric to stop the training process and select the best model for testing, as it has been commonly employed on object detection tasks [22,32,33]. The *mAP* can be calculated as follows:

$$mAP = \frac{1}{c} \sum_{i=1}^c AP_i, \quad (6)$$

where AP_i stands for the average precision value (for recall values from 0 to 1) of the i -th class.

We evaluated several metrics such as Root Mean Square Error (RMSE), Mean Squared Error (MSE), and MAE for the regression approaches. The best results were achieved using the MSE, which can be obtained by:

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (7)$$

5.2. Results

The results obtained by all approaches and models evaluated are shown in Table 2. In general, the detection-based models (e.g., YOLOv3 and YOLOv4) outperformed both regression and segmentation-free models for the ADMR task. Usually, complete systems (which perform detection and recognition using the same network) perform better than the sum of its isolated parts [40,56]. This is due to the fact that two-stage methods usually propagate the error from earlier stages through the entire pipeline. In addition, minimizing each loss separately is solving two different tasks, instead of jointly minimizing both losses, which is YOLO's target. Furthermore, segmentation-free approaches usually require a higher amount of data to train, as seen in [24,34].

Table 2

Reading results achieved on the UFPR-ADMR-v2 dataset.

Approach	Recog. rate (%)		MAE
	Meter	Dial	
Laroca et al. [24] (Multi-task [56])	15.30	64.24	24,181
Calefati et al. [44]	21.00	72.20	20,078
Gómez et al. [34]	23.50	70.56	24,074
Rosetta [51]	57.50	82.39	4,968
R ² AM [49]	59.60	83.72	4,087
STAR-Net [52]	60.00	83.70	9,454
GRCNN [48]	60.80	84.14	6,108
RARE [50]	61.80	84.25	5,696
CRNN [47]	64.20	86.27	3,669
TRBA [53]	66.80	87.87	2,404
Laroca et al. [24] (CR-NET [57])	71.40	91.94	1,680
ViTSTR-Base [45]	76.20	92.45	2,841
Laroca et al. [7]	78.30	94.07	1,327
AngReg (ResNet50-based [42])	85.00	96.38	429
AngReg (ResNet101-based [42])	86.20	96.52	402
AngReg (Xception-based [43])	88.00	96.96	244
YOLOv3 [33]	88.60	97.00	477
YOLOv4 [22]	91.10	97.67	242
YOLOv4 [22] (R)	92.00	98.00	197
YOLOv4 [22] (R) + AngReg (Xception-based [43])	92.50	98.11	187

The best results were achieved by combining the best object detection model (YOLOv4 [22]) with the best regression model (AngReg (Xception-based [43])). This generated our correction strategy, where AngReg indicated when the corrections should be performed, reducing the errors in the rightmost dials. We used the regression results to verify whether the dial needs to be corrected or not. It is noteworthy that this correction approach is only possible due to the approximate floating point annotated labels available in the UFPR-ADMR-v2 dataset. With an error tolerance of 1 kWh, our final system achieves an impressive MRR of 98.9%. In the next subsection, we detail why this error tolerance is acceptable. Samples of correctly predicted readings are shown in Fig. 7.

These results are significantly better than those achieved in our previous work [5]. The reasons are threefold: (i) there are now more than twice as many images for training the deep networks, which further highlights the importance of the UFPR-ADMR-v2 dataset; (ii) we fine-tuned several recent approaches and models for ADMR (as far as we know, there is no other work related to this problem with so many methods being evaluated); and (iii) we dealt specifically with each of the main causes of reading errors found in [5]: neighboring values, rotated images, lighting, and mirrored predictions, as described in the following subsection.

5.2.1. Error analysis

The main sources of errors in [5] were caused by:

- **Symmetry:** as there are clockwise and counterclockwise dials, in some cases the recognition methods cannot differentiate the direction and thus may output the mirrored value of the actual prediction.
- **Neighboring value:** the most common error. Variables such as angle, lighting, shadows and occlusion can impair the reading of a dial — especially when the dial is really close to the scale mark (almost changing from one digit to the next one). In addition, mechanical failures can lead to misaligned pointers causing *misalignment* errors (a subtype of neighboring value errors).
- **Severe lighting conditions & Dirt:** shadows, glares, reflections and dirt may confuse the networks, especially in low-contrast images, where those artifacts may emerge more than the pointer, fooling the network to think that it is a pointer border, resulting in an incorrect prediction.
- **Rotation:** tilted images are harder to predict, as the pointed value is not in the usual position. The predictions may be assigned to neighboring digits that would be in the current angle of the pointer if the image was not tilted.

² <https://github.com/AlexeyAB/darknet/>

³ <https://keras.io/>

⁴ <https://github.com/roatienza/deep-text-recognition-benchmark/>

Table 3
Distribution of errors by dial position.

Approach	Dial position				
	Frequency (%)				
	1	2	3	4	5
YOLOv3 [41]	7.86	8.57	7.14	11.43	65.00
YOLOv4 [22]	8.49	5.66	5.66	15.09	65.09
YOLOv4 [22] (R)	5.21	3.12	6.25	12.50	72.92
YOLOv4 [22] (R) + AngReg (Xception-based [43])	4.71	1.18	3.53	10.59	80.00

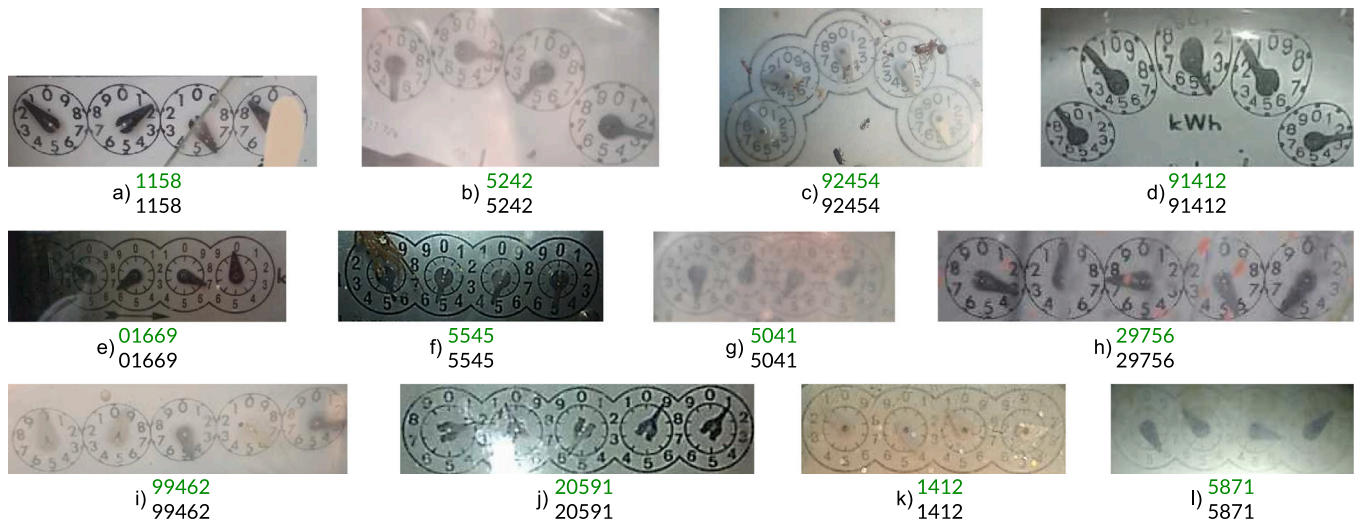


Fig. 7. Examples of correct readings performed by the approach that yielded the best results in our experiments: YOLOv4 (R) + AngReg (Xception).

The neighboring values issue was tackled by combining the regression with the detection approach. The regression approach uses the neighboring value to correctly yield a prediction, thus achieving a lower MAE even though the MRR and DRR are slightly worse than YOLOv3.

Symmetry- and lighting-related errors were addressed using various data augmentation techniques to create many images with different characteristics from a single labeled one. There was only one symmetry error made by the best model (i.e., YOLOv4 (R) + AngReg(Xception)).

As mentioned in Section 4.5, we solved the rotation issue using the angle between the first and last dial bounding boxes to rotate and rectify the dial image. This strategy improved the recognition rate achieved by YOLOv4 by 0.9%.

As listed in Table 3, most errors occurred on the least significant dial (i.e., the rightmost one). Although there are multi-dial errors, as shown in Fig. 9, these errors have MAE = 1. In fact, most of the errors (85%) have MAE = 1, as shown in Fig. 8. We estimate that such an error would only cost approximately R\$ 0.83 (US\$ 0.16) to the customer⁵ and would probably be compensated in the next monthly reading.

Moreover, it is hard even for humans to identify the correct reading due to perspective, lighting, shadows, and image quality. Indeed, there were disagreements between the authors regarding the correct label in some cases, which makes this type of error really hard to avoid without a policy of rounding up or down the reading value in such scenarios.

As previously mentioned, with an error tolerance of 1 kWh, our final system achieves an impressive MRR of 98.90%. Thus, we consider that the YOLOv4 (R) + AngReg (X) model can be reliably employed on real-world meter reading applications. However, there are still opportunities for enhancements since most of the errors occurred on images with inadequate quality and lighting, which should be double-checked by a reading specialist or rejected in a real-world system.

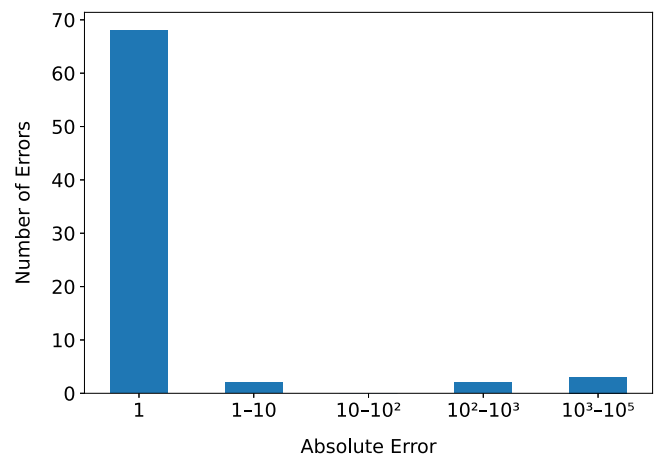


Fig. 8. Illustration of the number of errors per absolute error range. Note that most of the errors (85%) have MAE = 1. The number of errors refers to the 1000 images from the test set of UFPR-ADMR-v2.

Finally, the service company could easily detect the most significant errors using simple heuristic rules or even through a model capable of predicting the consumption of each consumer based on previous readings. If the reading is lower than the one recorded in the previous month or indicated that the consumer suddenly used a huge amount of energy, it can be double-checked by employees.

6. Conclusions

We tackled most of the issues regarding the automatic reading of dial meters. We came up with an approach that combines a detection model with a proposed new regression approach (AngReg) to enable error correction for the most significant dials, thus decreasing the MAE

⁵ <https://www.copel.com/hpcweb/copel-distribuicao/taxas-tarifas/>

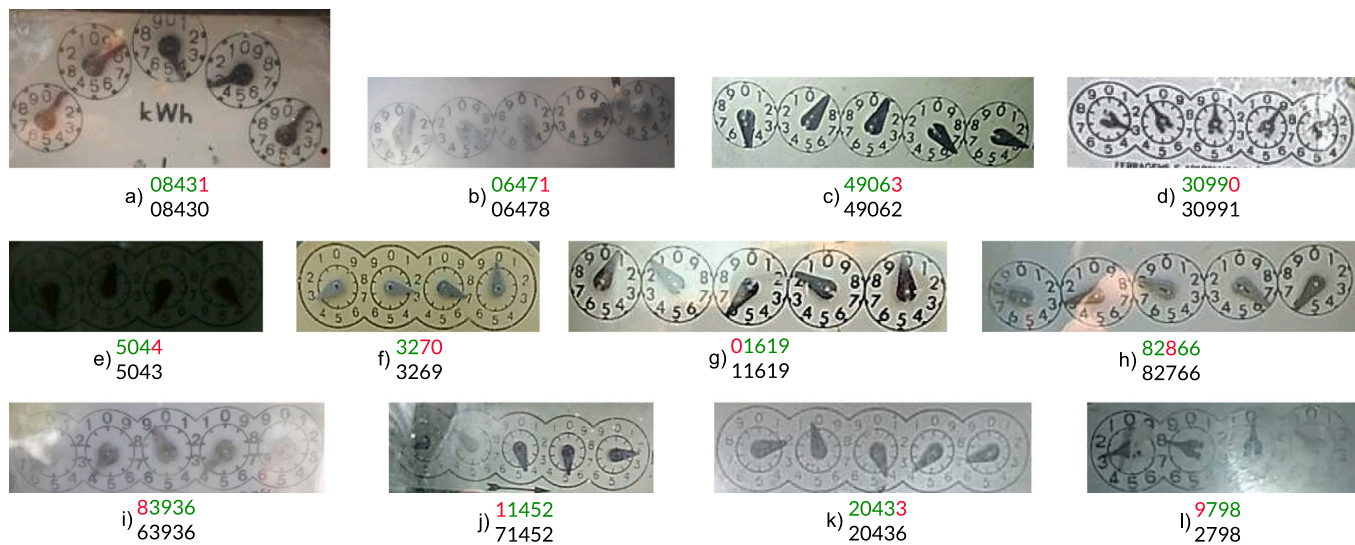


Fig. 9. Examples of reading errors made by the approach that achieved the best results in our experiments: YOLOv4 (R) + AngReg (Xception). The prediction (top reading) and the ground truth (bottom reading) are listed for each meter. The reading errors are outlined in red for better viewing. The error types are neighboring values (a, c, e, f), misalignment (g, h), severe lighting conditions (d, e, i, j), symmetry (k), and dirt (b, l).

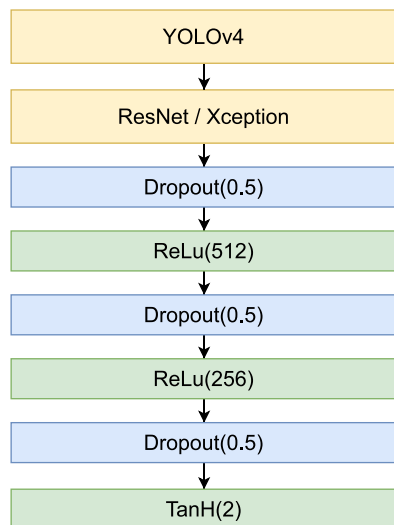


Fig. 10. Illustration of the AngReg topology. This architecture was obtained empirically by performing a grid search and optimizing hyperparameters.

considerably. We believe that such a robust approach can help reduce meter reading costs, reading errors, tampering and theft of meters, thus generating substantial savings for service companies.

More specifically, the neighboring values error – the most frequent type of error in [5] – was tackled using AngReg. Errors related to symmetry and lighting were addressed by creating many new training images through data augmentation. Lastly, we eliminated rotation errors by rotating the images according to the calculated angle between the detected dials.

We also introduced the UFPR-ADMR-v2 dataset, which is a public dataset for ADMR that includes 5000 fully-annotated meter images acquired in unconstrained scenarios by the service company's employees themselves. Compared to the UFPR-ADMR-v1 dataset, introduced in our previous work [5], the expanded version has more than twice the images and contains new annotations such as the approximated pointed value labels (with one decimal place precision) and segmentation masks for the dials. Such annotations enable the application of a much wider range of techniques for ADMR.

We also provided a detailed error analysis in order to assist the deployment of real-world ADMR systems. We showed that, by accepting minor errors (i.e., with MAE = 1) that have negligible cost (less than US\$0.16), we could increase the system's MRR from 92.50% to 98.90%. In fact, this cost would almost certainly be compensated in a subsequent reading.

As future work, we intend to develop a rejection system in order to automatically request new image samples in cases of poorly taken images. Additionally, we plan to leverage Generative Adversarial Networks (GANs) to generate more diverse training samples, ideally with pointers pointing equally at all angles (in terms of the number of dials across the images), to reduce bias by improving the balance between classes.

CRediT authorship contribution statement

Gabriel Salomon: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Rayson Laroca:** Conceptualization, Investigation, Writing – review & editing. **David Menotti:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in part by the National Council for Scientific and Technological Development (CNPq) (Grant 308879/2020-1), and in part by the Coordination for the Improvement of Higher Education Personnel (CAPES) (Grant 88887.516264/2020-00 and Social Demand Program). The Quadro RTX 8000 used for this research was donated by the NVIDIA Corporation. We also used another GPU (Titan V) donated by the NVIDIA Corporation through Professor André

Gregio's request. We gratefully acknowledge the Energy Company of Paraná (Copel), particularly the manager of the reading division Dihon Pereira Brandão, for providing the images for the creation of the UFPR-ADMR-v2 dataset.

Appendix A. AngReg topology

As can be inferred from Section 4.2 and Table 1, we implemented our AngReg approach using Darknet (YOLOv4) and Keras (ResNet/Xception).

More specifically, as detailed in Fig. 10, we explored YOLOv4 to first locate all dials in each input image.

Then, we employed the official Keras implementations of ResNet/Xception as an image feature extractor (without the top section). Finally, a custom FCN was implemented as the classifier, with MSE as the loss function.

Appendix B. Hyperparameters

In Darknet, we employed the same parameters as Laroca et al. [7]: Stochastic Gradient Descent (SGD) optimizer, 65K iterations, batch size = 64, and learning rate = $[10^{-3}, 10^{-4}, 10^{-5}]$ with decay steps at 26K and 45.5K. In Keras, we used the Adam optimizer, learning rate = 10^{-5} , batch size = 16 – 64 (depending on the model), max epochs = 100, and patience = 15 (patience refers to the number of epochs with no improvement after which training is stopped); these parameters were defined based on previous works [34,44] and also on preliminary experiments in the validation set. In PyTorch, as in [45,53], we adopted the following parameters: Adadelta optimizer (decay rate $\rho = 0.99$), 300K iterations, and batch size = 128.

References

- [1] M. Vanetti, I. Gallo, A. Nodari, GAS meter reading from real world images using a multi-net system, *Pattern Recognit. Lett.* 34 (5) (2013) 519–526.
- [2] I. Gallo, A. Zamberletti, L. Noce, Robust angle invariant GAS meter reading, in: *International Conference on Digital Image Computing: Techniques and Applications*, 2015, pp. 1–7.
- [3] C. Li, Y. Su, R. Yuan, D. Chu, J. Zhu, Light-weight spliced convolution network-based automatic water meter reading in smart city, *IEEE Access* 7 (2019) 174359–174367.
- [4] Y. Kabalci, A survey on smart metering and smart grid communication, *Renew. Sustain. Energy Rev.* 57 (2016) 302–318.
- [5] G. Salomon, R. Laroca, D. Menotti, Deep learning for image-based automatic dial meter reading: Dataset and baselines, in: *International Joint Conference on Neural Networks, IJCNN*, 2020, pp. 1–8.
- [6] U. S. Energy Information Administration, Electric power annual 2018, 2019, [Online]. Available: <https://www.eia.gov/electricity/annual/pdf/epa.pdf>.
- [7] R. Laroca, A.B. Araujo, L.A. Zanlorensi, E.C. de Almeida, D. Menotti, Towards image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach, *IEEE Access* 9 (2021) 67569–67584.
- [8] A. Naim, A. Aaroud, K. Akodadi, C. El Hachimi, A fully AI-based system to automate water meter data collection in Morocco country, *Array* 10 (2021) 100056.
- [9] A. Azeem, et al., A robust automatic meter reading system based on mask-RCNN, in: *IEEE International Conference on Advances in Electrical Engineering and Computer Applications*, 2020, pp. 209–213.
- [10] P. He, L. Zuo, C. Zhang, Z. Zhang, A value recognition algorithm for pointer meter based on improved mask-RCNN, in: *International Conference on Information Science and Technology*, 2019, pp. 108–113.
- [11] Y. Liu, J. Liu, Y. Ke, A detection and recognition system of pointer meters in substations based on computer vision, *Measurement* 152 (2020) 107333.
- [12] Y. Huang, X. Dai, Q. Meng, An automatic detection and recognition method for pointer-type meters in natural gas stations, in: *Chinese Control Conference*, 2019, pp. 7866–7871.
- [13] W. Cai, B. Ma, L. Zhang, Y. Han, A pointer meter recognition method based on virtual sample generation technology, *Measurement* 163 (2020) 107962.
- [14] B. Howells, J. Charles, R. Cipolla, Real-time analogue gauge transcription on mobile phone, in: *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021, pp. 2369–2377.
- [15] Y. Fang, Y. Dai, G. He, D. Qi, A mask RCNN based automatic reading method for pointer meter, in: *Chinese Control Conference*, 2019, pp. 8466–8471.
- [16] H. Jiale, L. En, T. Bingjie, L. Ming, Reading recognition method of analog measuring instruments based on improved hough transform, in: *International Conference on Electronic Measurement Instruments*, Vol. 3, 2011, pp. 337–340.
- [17] L. Zhang, et al., Pointer-type meter automatic reading from complex environment based on visual saliency, in: *International Conference on Wavelet Analysis and Pattern Recognition*, 2016, pp. 264–269.
- [18] W. Zheng, et al., Development of an automatic reading method and software for pointer instruments, in: *International Conference on Electronics Instrumentation Information Systems*, 2017, pp. 1–6.
- [19] R. Ocampo-Vega, et al., Image processing for automatic reading of electro-mechanical utility meters, in: *Mexican International Conference on Artificial Intelligence*, 2013, pp. 164–170.
- [20] R.C. da Silva Marques, A.C. Serra, J.V.F. França, J.O.B. Diniz, G. Braz, J.D.S. de Almeida, M.I.A. da Silva, E.M.G. Monteiro, Image-based electric consumption recognition via multi-task learning, in: *Brazilian Conference on Intelligent Systems*, 2019, pp. 419–424.
- [21] ANEEL – Agência Nacional de Energia Elétrica, Perdas de energia elétrica na distribuição, 2019, [Online]. Available: <https://www.aneel.gov.br/documents/654800/18766993/Relat%C3%B3rio+Perdas+de+Energia+2019.pdf>.
- [22] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, YOLOv4: Optimal speed and accuracy of object detection, 2020, pp. 1–14, arXiv preprint arXiv:2004.10934.
- [23] Y. Tang, C. Ten, C. Wang, G. Parker, Extraction of energy information from analog meters using image processing, *IEEE Trans. Smart Grid* 6 (4) (2015) 2032–2040.
- [24] R. Laroca, V. Barroso, M.A. Diniz, G.R. Gonçalves, W.R. Schwartz, D. Menotti, Convolutional neural networks for automatic meter reading, *J. Electron. Imaging* 28 (1) (2019) 013023.
- [25] C. Tsai, T.D. Shou, S. Chen, J. Hsieh, Use SSD to detect the digital region in electricity meter, in: *International Conference on Machine Learning and Cybernetics, ICMLC*, 2019, pp. 1–7.
- [26] F. Yang, L. Jin, S. Lai, X. Gao, Z. Li, Fully convolutional sequence recognition network for water meter number reading, *IEEE Access* 7 (2019) 11679–11687.
- [27] L. Hou, H. Qu, Automatic recognition system of pointer meters based on lightweight CNN and WSNs with on-sensor image processing, *Measurement* 183 (2021) 109819.
- [28] C. Zheng, S. Wang, Y. Zhang, P. Zhang, Y. Zhao, A robust and automatic recognition system of analog instruments in power system by using computer vision, *Measurement* 92 (2016) 413–420.
- [29] E. Luz, et al., Deep pericocular representation aiming video surveillance, *Pattern Recognit. Lett.* 114 (2020) 2–12.
- [30] L. Zuo, P. He, C. Zhang, Z. Zhang, A robust approach to reading recognition of pointer meters based on improved mask-RCNN, *Neurocomputing* 388 (2020) 90–101.
- [31] R. Sowah, et al., An intelligent instrument reader: Using computer vision and machine learning to automate meter reading, *IEEE Ind. Appl. Mag.* (2021).
- [32] S. Ren, et al., Faster r-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6) (2017) 1137–1149.
- [33] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, 2018, arXiv preprint arXiv:1804.02767.
- [34] L. Gómez, et al., Cutting Sayre's knot: Reading scene text without segmentation. Application to utility meters, in: *IAPR International Workshop on Document Analysis Systems*, 2018, pp. 97–102.
- [35] J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2017, pp. 6517–6525.
- [36] A. Nodari, I. Gallo, A multi-neural network approach to image detection and segmentation of gas meter counter, in: *IAPR Conference on Machine Vision Applications, MVA*, 2011, pp. 239–242.
- [37] A. Iqbal, A. Basit, I. Ali, J. Babar, I. Ullah, Automated meter reading detection using inception with single shot multi-box detector, *Intell. Autom. Soft Comput.* 27 (2021) 299–309.
- [38] Copel, Energy company of Paraná, 2021, [Online]. Available: <http://www.copel.com/hpcopel/english/>.
- [39] X. Kou, S. Liu, K. Cheng, Y. Qian, Development of a YOLO-V3-based model for detecting defects on steel strip surface, *Measurement* 182 (2021) 109454.
- [40] R. Laroca, L.A. Zanlorensi, G.R. Gonçalves, E. Todt, W.R. Schwartz, D. Menotti, An efficient and layout-independent automatic license plate recognition system based on the YOLO detector, *IET Intell. Transp. Syst.* 15 (4) (2021) 483–503.
- [41] S. Liao, P. Zhou, L. Wang, S. Su, Reading digital numbers of water meter with deep learning based object detector, in: *Chinese Conference on Pattern Recognition and Computer Vision*, 2019, pp. 38–49.
- [42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2016, pp. 770–778.
- [43] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2017, pp. 1251–1258.
- [44] A. Calefati, I. Gallo, S. Nawaz, Reading meter numbers in the wild, in: *Digital Image Computing: Techniques and Applications, DICTA*, 2019, pp. 1–6.
- [45] R. Atienza, Vision transformer for fast and efficient scene text recognition, in: *International Conference on Document Analysis and Recognition, ICDAR*, 2021.

- [46] R. Laroca, E.V. Cardoso, D.R. Lucio, V. Estevam, D. Menotti, On the cross-dataset generalization in license plate recognition, in: International Conference on Computer Vision Theory and Applications, VISAPP, 2022, pp. 166–178.
- [47] B. Shi, X. Bai, C. Yao, An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (11) (2017) 2298–2304.
- [48] J. Wang, X. Hu, Gated recurrent convolution neural network for OCR, in: Conference on Neural Information Processing Systems, 2017.
- [49] C. Lee, S. Osindero, Recursive recurrent nets with attention modeling for OCR in the wild, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 2231–2239.
- [50] B. Shi, X. Wang, P. Lyu, C. Yao, X. Bai, Robust scene text recognition with automatic rectification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4168–4176.
- [51] F. Borisyuk, A. Gordo, V. Sivakumar, Rosetta: Large scale system for text detection and recognition in images, in: ACM SIGKDD Intl. Conference on Knowledge Discovery & Data Mining, 2018, pp. 71–79.
- [52] W. Liu, C. Chen, K.-Y.K. Wong, Z. Su, J. Han, STAR-Net: A SpaTial attention residue network for scene text recognition, in: British Machine Vision Conference, BMVC, 2016, pp. 1–13.
- [53] J. Baek, et al., What is wrong with scene text recognition model comparisons? Dataset and model analysis, in: IEEE/CVF International Conference on Computer Vision, ICCV, 2019, pp. 4714–4722.
- [54] C. Yang, W. Xie, A. Zisserman, It's about time: Analog clock reading in the wild, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 1–16.
- [55] A. Buslaev, V.I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, A.A. Kalinin, Albumentations: Fast and flexible image augmentations, *Information* 11 (2) (2020).
- [56] G.R. Gonçalves, et al., Multi-task learning for low-resolution license plate recognition, in: Iberoamerican Congress on Pattern Recognition, CIARP, 2019, pp. 251–261.
- [57] S.M. Silva, C.R. Jung, Real-time license plate detection and recognition using deep convolutional neural networks, *J. Vis. Commun. Image Represent.* (2020) 102773.