

UNIVERSIDADE FEDERAL DO PARANÁ

GABRIEL SALOMON ANICETO

AUTOMATIC DIAL METER READING

CURITIBA PR

2020

GABRIEL SALOMON ANICETO

AUTOMATIC DIAL METER READING

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: David Menotti.

CURITIBA PR

2020

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

A597a Aniceto, Gabriel Salomon
Automatic dial meter reading [recurso eletrônico] / Gabriel
Salomon Aniceto – Curitiba, 2020.

Dissertação - Universidade Federal do Paraná, Setor de
Ciências Exatas, Programa de Pós-Graduação em Informática.

Orientador: David Menotti Gomes

1. Leitura de medidores (Energia elétrica). 2. Medidores de
relógio - Leitura automática. I. Universidade Federal do Paraná.
II. Gomes, David Menotti. III. Título.

CDD: 006.24

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **GABRIEL SALOMON ANICETO** intitulada: **Automatic Dial Meter Reading**, sob orientação do Prof. Dr. DAVID MENOTTI GOMES, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 24 de Setembro de 2021.

Assinatura Eletrônica

27/09/2021 10:14:44.0

DAVID MENOTTI GOMES

Presidente da Banca Examinadora

Assinatura Eletrônica

27/09/2021 10:10:22.0

PAULO RICARDO LISBOA DE ALMEIDA

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

30/09/2021 15:17:46.0

CLAUDIO ROSITO JUNG

Avaliador Externo (UNIVER. FEDERAL DO RIO GRANDE DO SUL)

Assinatura Eletrônica

27/09/2021 13:35:57.0

LUIZ EDUARDO SOARES DE OLIVEIRA

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

*“Who wants to live forever?”
Freddie Mercury*

ACKNOWLEDGEMENTS

I would like to thank my advisor, David Menotti, my lab colleagues, specially Rayson Laroca that helped with some of the experiments and considerations.

I would also like to thank the members of my dissertation committee, Prof. Eduardo Todt and Prof. Luiz Oliveira, for the time they spent working on my committee, and for their great help in improving the quality of this work.

I would also like to thank my family for the immense support, both emotionally and financially. My parents are incredible people, without them I would not be always in the pursue of knowledge and I would not have the notions of moral values and ethics that I have today.

This work was supported by the National Council for Scientific and Technological Development (CNPq) and also by the Coordination for the Improvement of Higher Education Personnel (CAPES).

I gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

I also thank the Energy Company of Paraná (Copel) for providing the images for the UFPR-ADMR-v2 dataset.

RESUMO

Houve muitos avanços no campo da leitura automática de medidores. Uma das mais significativas foi a introdução de medidores inteligentes — medidores que podem realizar e enviar automaticamente a leitura para o provedor de serviço. No entanto, os medidores inteligentes têm a desvantagem de depender de cabos de dados ou redes móveis para realizar as leituras. Estimamos que a Companhia Paranaense de Energia (Copel), no Brasil, realiza mais de 4 milhões de leituras de medidores por mês, sendo 850 mil leituras de medidores de relógio (ponteiro). Portanto, um sistema de leitura automática baseado em imagens poderia agilizar o processo de leitura, reduzir erros humanos, criar uma prova de leitura e, por fim, permitir que os próprios clientes façam a leitura por meio de um aplicativo móvel. Nossas principais contribuições são: (i) novas abordagens para a leitura automática do medidor de relógio (ADMR); (ii) o maior (até onde sabemos) conjunto de dados de medidores de relógio do mundo real disponível publicamente (compartilhado mediante solicitação), chamado UFPR-ADMR-v2; (iii) uma avaliação de várias abordagens diferentes, incluindo métodos de última geração para realizar ADMR. Nosso método de melhor desempenho foi a combinação de YOLOv4 (rede de detecção) e Xception (treinada para regressão), com várias técnicas de pós-processamento. Em comparação com nosso trabalho anterior, fomos capazes de evoluir de 17,66% dos erros concentrados nos relógios menos significativos para 80%. O Erro Quadrado Médio diminuiu de 1.343 para 129. Com nossas técnicas e abordagens de correção, conseguimos atingir uma Taxa de Reconhecimento de Medidores (MRR) de 98,90 % (com uma tolerância de erro de 1 kWh).

Palavras-chave: Leitura automática de medidores de relógio. Leitura de medidores. Aprendizado profundo.

ABSTRACT

There has been a lot of advances in the field of Automatic Meter Reading. One of the the most significant was the introduction of smart meters — meters that can automatically perform and send the reading to the service provider. Nevertheless, smart meters have the disadvantage of depending on data cables or mobile networks to perform the readings. We estimate that the Energy Company of Paraná (Copel), in Brazil, performs more than 4 million meter readings per month, being 850 thousand of them dial meter readings. Therefore, an image-based automatic reading system could speed up the reading process, reduce human errors, create a proof of reading and ultimately enable the customers to perform the reading themselves through a mobile application. Our main contributions are: (i) novel approaches towards Automatic Dial Meter Reading (ADMR); (ii) the largest (to the best of our knowledge) publicly available real-world dial meter dataset (shared upon request) called UFPR-ADMR-v2; (iii) an evaluation of several different approaches including state-of-the-art methods towards performing ADMR. Our best performing method was the combination of YOLOv4 (detection) and Xception (regression), with several post-processing techniques. Compared to our previous work, we were able to evolve from 17.66% of the errors concentrated on the least significant dial to 80%. The Mean Absolute Error decreased from 1,343 to 129. With our correction techniques and approaches, we were able to achieve a Meter Recognition Rate of 98.90% (with an error tolerance of 1kWh).

Keywords: Automatic dial meter reading. Meter reading. Deep learning.

LIST OF FIGURES

1.1	The most common types of residential energy meters.	13
2.1	The most common internal components of a gauge (single dial meter). Adapted from Lauridsen et al. (2019).	17
2.2	Illustration of the energy meter dials, note the alternating direction between sequential dials, being the least significant (rightmost) dial always oriented clockwise.	18
2.3	The IoU is the ratio between the overlapping area of the bounding boxes and the area of their union. Extracted from https://www.pyimagesearch.com/	18
2.4	ResNet's residual block. The shortcut path (on the right) introduces lower level features directly to higher abstraction layers, speeding up the learning process and preserving information. Extracted from He et al. (2016).	21
2.5	The Modified Depthwise Separable Convolution consists in first applying a pointwise convolution followed by a depthwise (channel-wise convolution). Extracted from Towards Data Science ¹	22
2.6	Illustration of the $S \times S$ grid. For each grid cell, B bounding boxes are generated, with their respective confidence. Besides, the class probability map illustrates the predominant class for each cell. Extracted from Redmon et al. (2016).	23
3.1	A common single dial (gauge) recognition pipeline. Extracted from Zhang et al. (2016).	24
3.2	Comparison between HCT, LS and PSO. Adapted from Zhu et al. (2019).	25
3.3	Perspective correction achieved using a transformation matrix. Extracted from Liu et al. (2019b).	26
3.4	Illustration of the annotated points: one point for each scale mark, and five points for the pointer. Extracted from Fang et al. (2019).	26
3.5	Pipeline for automatic gauge reading. The dial and pointer segmentation were performed with Mask R-CNN. After perspective correction, the pointed value is recognized using PCA and the gauge recognized parameters. Extracted from He et al. (2019).	27
3.6	To perform the dial reading, the two points (centroid and the furthest point from it) were chosen, and a line defined by them is used for the angle calculation (the angle between the line and the x-axis). Extracted from Ocampo-Vega et al. (2013).	27
3.7	Images were captured every 15 minutes and send wirelessly to a remote server. Extracted from Tang et al. (2015).	28
3.8	Illustration of each step of the meter recognition process. Extracted from Vanetti et al. (2012).	29
3.9	The shallow segmentation-free network. Extracted from Gómez et al. (2018).	30
3.10	Electronic meter images. Extracted from the proposed dataset Kanagarathinam and Sekar (2019).	30

4.1	Examples of the images chosen for the dataset. Note that the costumer meter identification is blurred for keeping subject privacy.	32
4.2	The labeling process with our custom made annotation tool consists of the following steps: (a) estimating the four corners of the irregular bounding rectangle; (b) rectifying the rectangle for annotation of ellipses; (c) generating the final bounding boxes in the original (non-rectified) image around each ellipse, together with the irregular bounding rectangle.. . . .	33
4.3	The distribution of digits according to the dial position on the meters from the UFPR-ADMR dataset. As 51.8% of the meters in the proposed dataset do not have a 5th dial, the values of the 5th dial quantities were interpolated for better visualization.	34
4.4	The distribution of digits according to the dial position on the meter, on the digit-based meters from UFPR-AMR dataset (Laroca et al., 2019).	34
4.5	Distribution of the pointer positions in the UFPR-ADMR-v2 dataset.	35
4.6	Samples of the challenging scenarios present on the provided images. We selected for the UFPR-ADMR-v2 dataset 5,000 images in which it was possible for a human to recognize the correct reading of the meter. Note that most of the dials presented above are not readable, and were not selected for the dataset. We blurred the region containing the consumer unit number in each image due to privacy constraints. None of the images above were selected to our database, they are displayed solely for illustration purposes.	36
5.1	Proposed pipelines.	37
5.2	Illustration of the angle correction method. The angle (θ) between the line that passes through the middle of the first and last detected dial and the x -axis is computed. Afterwards, the correction method rotates the image $-\theta$ degrees in order to make the dials parallel to the x -axis.	40
6.1	Prediction (top number) and ground-truth (bottom number) examples of correctly recognized meters (the images were cropped for better visualization).	43
6.2	Prediction (top number) and ground-truth (bottom number) examples (the images were cropped for better visualization) of the errors of the best method (YOLOv4 (R) + Xception. The errors are marked in red and include: neighbor values (a,b,c,e,h,i,m,o); symmetry (d); misalignment (f,p); severe lighting conditions (g,j,k,n); and dirt(l).	45
6.3	For each approach two samples are given (one per line). For each sample: on the left the masks are displayed, on the right the original dial (for reference) and in the middle the synthetic images generated by Pix2Pix.	47
6.4	Samples of the 50,000 generated images. After generating the individual dials, using the annotated mask and bounding rectangle, we correct the perspective and insert the synthetic dials in the original image. Afterwards, the image is warped back into the original perspective.	47

LIST OF TABLES

3.1	Summary of the reviewed works.....	31
4.1	Statistics about the size of meters and individual dials.	33
6.1	Dial detection results achieved on the UFPR-ADMR-v2 dataset.	42
6.2	Reading results achieved on the UFPR-ADMR-v2 dataset.....	43
6.3	Distribution of errors by dial position.	45
6.4	The trade-off between accuracy and cost of error using YOLOv4 (Rotated) + Xception and the number of manual inspections required to perform all correct readings in the testing dataset (1,000 dials images).....	45

LIST OF ACRONYMS

ADMR	Automatic Dial Meter Reading
AMR	Automatic Meter Reading
ANEEL	Agência Nacional de Energia Elétrica
CNN	Convolutional Neural Network
Copel	Energy Company of Paraná
CRNN	Convolutional Recurrent Neural Network
DRR	Dial Recognition Rate
FCSRNN	Fully Convolutional Sequence Recognition Network
FPS	Frames Per Second
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HCT	Hough Circle Transform
HLT	Hough Line Transform
HOG	Histogram of Oriented Gradients
IoU	Intersection over Union
kNN	k-Nearest Neighbors
LPIF	Localized Pixel Inhomogeneity Factor
LS	Least Squares
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
mAP	mean Average Precision
MLP	Multilayer Perceptron
MRR	Meter Recognition Rate
MSE	Mean Squared Error
MSER	Maximally Stable Extremal Regions
NMS	Non-maximum Suppression
OCR	Optical Character Recognition
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
PTZ	Pan-tilt-zoom
R-CNN	Region-based Convolutional Neural Network
R-FCN	Region-based Fully Convolutional Network
RBFNN	Radial Basis Function Neural Network
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
RoI	Region of Interest
RPN	Region Proposal Network
SIFT	Scale-Invariant Feature Transform
SSD	Single Shot MultiBox Detector
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
YOLO	You Only Look Once

CONTENTS

1 INTRODUCTION	13
1.1 MOTIVATION	14
1.2 OBJECTIVES	15
1.3 CONTRIBUTIONS	16
1.4 ROADMAP	16
2 THEORETICAL BACKGROUND	17
2.1 DIAL METERS COMPONENTS	17
2.2 ENERGY DIAL METERS DESIGN	17
2.3 DETECTION METRICS	18
2.4 ADMR METRICS	20
2.5 RESIDUAL NETWORKS	21
2.6 YOLO	22
2.7 SEGMENTATION-FREE NETWORKS	23
3 RELATED WORK	24
3.1 SINGLE DIAL METERS	24
3.2 MULTI-DIAL METERS	27
3.3 DIGIT-BASED METERS	28
3.4 DATASETS	30
3.5 OVERVIEW	31
4 THE UFPR-ADMR-V2 DATASET	32
4.1 CHALLENGES	35
5 PROPOSED APPROACH	37
5.1 OBJECT DETECTION WITH RECOGNITION	37
5.2 OBJECT DETECTION WITH REGRESSION	38
5.3 SEGMENTATION-FREE RECOGNITION	39
5.4 DATA AUGMENTATION	39
5.5 POST-CORRECTION	40
6 EXPERIMENTAL RESULTS	41
6.1 TRAINING ENVIRONMENT AND EVALUATION METRICS	41
6.2 EVALUATION PROTOCOL	41
6.3 RESULTS	42
6.4 ERROR ANALYSIS	44
6.5 GAN GENERATED DATASET	46

7 CONCLUSIONS	48
7.1 FUTURE WORK	48
7.2 PUBLICATIONS	49
REFERENCES	50

1 INTRODUCTION

Measuring residential energy consumption is known to be a laborious task (Vanetti et al., 2012; Gallo et al., 2015; Li et al., 2019). Although smart meters are gradually replacing old meters, there are still many old mechanical meters in operation around the world since their replacement is time-consuming and costly. In many regions, such as remote areas and developing nations, manual on-site readings are still prevalent (Laroca et al., 2019). Even in developed countries, replacements are still far from complete. For example, in the end of 2018, there were still more than 26 million non-automatic meters in the United States (U.S. Energy Information Administration, 2019) representing 17% of the total meters (it's unclear however the proportion of each sub-type of analog meters, e.g. dial meters, cyclometers).

In the literature, the task of performing meter readings with technological assistance is denoted Automatic Meter Reading (AMR) and is usually associated with digital and smart meters (Kabalcı, 2016). We use this designation exclusively for image-based automatic readings. AMR allows the employees of the service company (electricity/gas/water) or, preferably, the consumers themselves to capture meter images using a mobile device, which is cheaper and more feasible than manual on-site reading, and easier to deploy – in the short/medium term – than the replacement of old meters.

There are two main categories of residential energy meters (Ausgrid, 2020; Callmepower, 2020): (i) analog (with cyclometer and dial displays) and (ii) digital (with electronic display and smart meters), as shown in Figure 1.1.

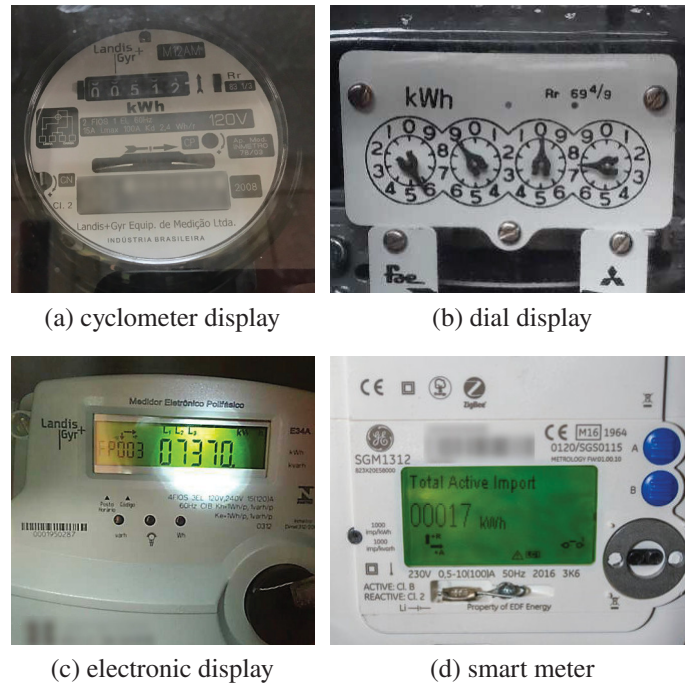


Figure 1.1: The most common types of residential energy meters.

This work focuses on Automatic Dial Meter Reading (ADMR). Even though there are numerous energy dial meters still on operation, the subject has not been widely explored in the literature. Our main assumption regarding the scarcity of ADMR literature relies upon the few number of gas/energy/water companies and the advent of smart meters. Furthermore, the

techniques used previously to tackle this type of problem were not robust to many variations (Zheng et al., 2017; Huang et al., 2019), limiting their deployment to well-controlled indoor environments. Additionally, many of the real implemented systems must be intellectual property of the service companies, thus not available for the general public.

Most of the existing dial meter recognition literature is focused on single dial meters (a.k.a. gauges or single-pointer meters) deployed for industrial applications, e.g., pressure meters (Liu et al., 2019b; Zheng et al., 2017; Huang et al., 2019), voltmeter (Jiale et al., 2011) and ammeter (Fang et al., 2019). Single-pointer meter reading is imperative to avoid accidents in many industrial facilities. According to Gao et al. (2018), humans are prone to misreading, fake the readings (to avoid labor), so many substations use inspection robots for monitoring crucial equipment (Gao et al., 2018). The robots are also useful in case of gas leaks, which restrains human access to the location for inspection.

A key difference between the gauge literature and ADMR arises from the fact that in most industrial applications, the meters are indoors, in a controlled environment, which is not the same as energy meters in outdoor environments. Outdoor meters normally present challenging conditions, e.g., severe lighting conditions (low light, glares, uneven illumination, reflections, etc.), dirt in the region of interest, meter degradation (erosion caused by the weather) and images taken at a distance.

In addition, most gauge approaches are based on handcrafted features (Tang et al., 2015; Zheng et al., 2017), and were evaluated exclusively on private datasets (Ocampo-Vega et al., 2013; Tang et al., 2015; Zheng et al., 2017; Fang et al., 2019; Liu et al., 2019b).

Another challenge, regarding methods comparison is the lack of detailed algorithms and procedures, as stated by Lauridsen et al. (2019). Some of the works only mention the techniques or methods but do not provide parameters or the algorithm to enable replication (Ren et al., 2019; Tang et al., 2015; Zhao et al., 2005a). Besides, most works are self-contained, not providing comparison with other methods in the literature (Souare et al., 2014).

Considering the aforementioned discussions, in this project we introduce a real-world fully-labeled dataset (shared under request) containing 5,000 meter images, acquired in unconstrained scenarios by the Energy Company of Paraná (Copel) employees, with 22,410 individual dials and a well-defined evaluation protocol to assist the development and assessment of new approaches for this task¹. To the best of our knowledge, there are no public datasets containing multi dial meter images, nor dial energy meter datasets available in the literature.

We conducted experiments using deep learning models in our dataset images to serve as baselines for future work, investigating problems related to dial meter reading and providing guidance for further research through a detailed quantitative and qualitative error analysis. As far as we know, this is the first dial meter recognition work to propose a completely deep learning-based pipeline.

1.1 MOTIVATION

Copel (2020) measures electricity consumption in more than 4 million consuming units (i.e., meters) per month in the Brazilian state of Paraná. And, from the images they provided to us (see Section 4), we estimate that 21% of these devices are dial meters, resulting in more than 840,000 monthly dial meter readings, only in the state of Paraná.

Considering electricity losses (da Silva Marques et al., 2019), according to the ANEEL’s (Brazil’s National Electricity Agency) Annual Report of 2019 (ANEEL, 2019), approximately

¹The UFPR-ADMR-v2 dataset is **publicly available** (under request) to the research community at web.inf.ufpr.br/vri/databases/ufpr-admr/.

46% of the electricity losses in Brazil are non-technical and occur due to causes such as theft, tampering with meters, non-working meters, reading and billing errors. However, the proportion of occurrence for the aforementioned events is not informed in the report. In absolute values, the non-technical losses represent 33.3TWh lost per year (6.6% of the generated electricity).

A robust image-based AMR system could speed up the reading process, reduce human errors, create a proof of reading to avoid theft and detect non-working meters and ultimately enable the customers to perform the reading themselves through a mobile application. Hence, reducing the service provider's reading cost and reducing the requirement of a monthly visit of the service provider's workers to perform the reading.

In fact, recent regulations promulgated by ANEEL authorizes self-reading for any consumer². Those regulations enable the service companies to provide a self-reading alternative for the customers, specially for those located in difficult to reach areas, promoting the reduction of on-site readings frequency. An ADMR system could enable the consumers to perform the reading by themselves, avoiding frauds and after-hours readings (for consumers which are not at home during business hours to allow access to hard-to-reach meter reading).

The introduction of a novel challenging dataset arises from the fact that many works in the literature propose methods that work well in controlled images (Tang et al., 2015; Zheng et al., 2017), but in a real-world scenario an AMR system should overcome quality issues to perform well. We faced several issues regarding image acquisition, such as glass reflections, dirt, broken meter glass, and also regarding image quality, such as noisy, blurred, low-lighting and low-contrast images. In addition, we have the occurrence of misaligned dials (usually caused by manufacturing defects and/or weathering) which require us to analyze the next dial (if available) in order to yield a correct prediction.

In our work, we explore distinct methods such as regression and segmentation-free for their advantages in comparison with detection methods. Detection methods usually produce discrete labels, which can not be the most suitable to correct predictions in sequences of dials, as they do not contain any information about the exact location of pointers, just an integer. Regression methods can learn more precise labels (regarding decimal places) and segmentation-free methods take into account the spatial information; thus, theoretically being capable of using the other dials to yield a more precise prediction (if a sufficient amount of data is provided).

1.2 OBJECTIVES

Our main objective is to introduce efficient new approaches towards ADMR, with a complete pipeline, from image to reading. Efficiency can be evaluated using the recognition rate and absolute error. In order to achieve the main objective, several specific objectives need to be achieved:

- Obtain/Annotate 5,000 realistic images of dial meters for evaluation of the proposed system;
- Investigate the approaches proposed in the literature, identifying their advantages and limitations;
- Propose new approaches towards efficiently solving the problem of ADMR;
- Use/Propose metrics of evaluation, along with a protocol, for the sake of reproducibility and fair comparison;
- Perform an error analysis of the prediction errors in order to improve the system.

²https://www.aneel.gov.br/sala-de-imprensa-exibicao-2/-/asset_publisher/zXQREz8EVIZ6/content/aneel-aprova-novas-regras-para-leitura-e-medicao-de-energia-eletrica/656877/

1.3 CONTRIBUTIONS

Our main contributions are:

- A new publicly available dataset (under request) consisting of 5,000 labeled images with 22,410 annotated dials, along with segmentation masks and integer and rational labels for each dial;
- Novel approaches towards solving the issue of ADMR using detection, regression and segmentation-free networks;
- New metrics for fair ADMR methods evaluation, under distinct performance aspects;
- An error analysis of the main issues that can degrade performance in ADMR systems;
- A robust system to perform ADMR in unconstrained scenarios, achieving an Meter Recognition Rate (MRR) of 98.90% (with an error tolerance of 1kWh).

1.4 ROADMAP

The remainder of this work is organized as follows. The theoretical background is presented in Chapter 2. In Chapter 3, we discuss approaches designed for AMR. The proposed dataset is described in Chapter 4. Chapter 5 presents the evaluated deep learning-based approaches for automatic reading of dial meters, while the results (with a detailed error analysis) are reported in Chapter 6. Lastly, in Chapter 7, we state the conclusions and future work.

2 THEORETICAL BACKGROUND

In this chapter, we introduce theoretical concepts regarding the dial meters' main components, evaluation metrics (detection and ADMR metrics), regression, residual networks, Faster R-CNN and YOLO. As it is beyond the scope of this work, basic concepts regarding deep learning, Convolutional Neural Networks (CNNs) and neural networks are not presented. We strongly recommend the book *Deep Learning* (Goodfellow et al., 2016) for more information on such topics.

2.1 DIAL METERS COMPONENTS

The main components of a regular gauge (single dial meter) are: pointer, dial, dial number and scale mark (Lauridsen et al., 2019), as shown in Figure 4.1. Unfortunately, no equivalent detailed illustration on the dial energy meter literature was located, so we extend the pressure gauge's definitions for the dial energy meters.

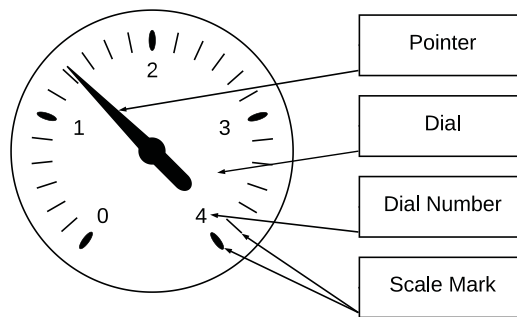


Figure 2.1: The most common internal components of a gauge (single dial meter). Adapted from Lauridsen et al. (2019).

2.2 ENERGY DIAL METERS DESIGN

The energy dial meters are usually composed of 4 or 5 dials, depending on the model. There is a pattern in the dial direction design (clockwise vs. counterclockwise) during the annotation process: the dials have alternating directions, with the least significant dial always clockwise oriented (CW). 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 we can see in Fig 2.2. This information is useful to train methods focused on the angle of the pointer regardless of the orientation. The orientation can then be retrieved using the number of dials detected and their index on the meter.

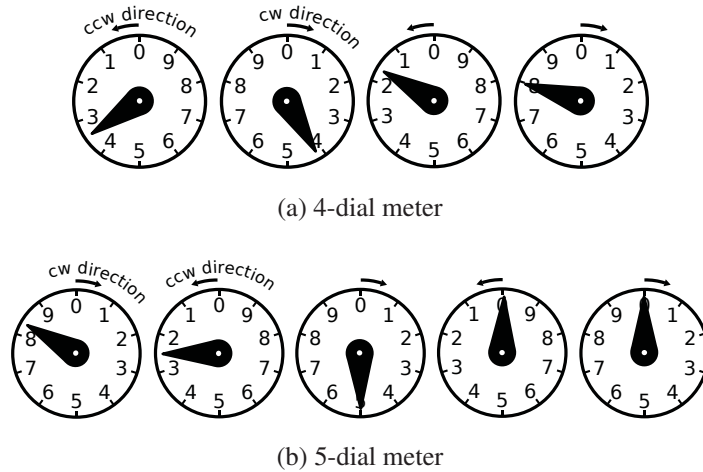


Figure 2.2: Illustration of the energy meter dials, note the alternating direction between sequential dials, being the least significant (rightmost) dial always oriented clockwise.

2.3 DETECTION METRICS

As we use detection networks to detect the dials, some common concepts of object detection are introduced below, such as Intersection over Union (IoU), mean Average Precision (mAP). IoU is a common metric in object detection, for measuring the accuracy of the object detection. It is calculated using the ratio between the area where the predicted bounding box and the ground truth bounding box intersect, and the area of their union (see Figure 2.3). The formulation below can be used to obtain the metric:

$$IoU = \frac{area(pred \cap gt)}{area(pred \cup gt)}. \quad (2.1)$$

The IoU between a predicted bounding box and the ground truth can determine (using a defined threshold) if an object was detected, and additionally how accurate is the detection. This metric is also used to perform Non-maximum Suppression (NMS). NMS eliminates multiple overlapping detections of the same object in different scales and sizes, keeping only the detection with the higher confidence (as it is probably the most accurate).

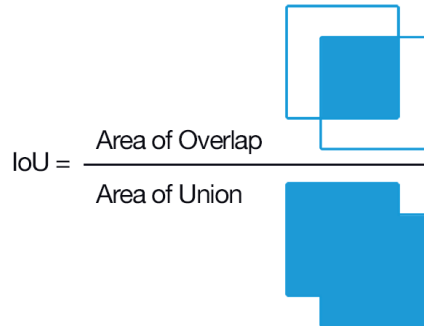


Figure 2.3: The IoU is the ratio between the overlapping area of the bounding boxes and the area of their union. Extracted from <https://www.pyimagesearch.com/>.

We use in the training stage (to perform validation), mAP, a commonly employed metric for object detection networks (Ren et al., 2017; Redmon et al., 2016; Redmon and Farhadi, 2018).

Before explaining mAP, we should first introduce the concepts of precision and recall. Precision and Recall are formally defined as follows:

$$\begin{aligned} Precision &= \frac{TP}{TP + FP}, \\ Recall &= \frac{TP}{TP + FN}. \end{aligned} \quad (2.2)$$

where TP (True Positives) are the positive samples that are correctly classified, FP (False Positives) are the negative samples incorrectly classified as positive and FN (False Negatives) represent the samples that were classified as negative, but are actually positive.

Intuitively, precision measures if all the samples classified as positive are indeed positive. Recall measures if there are no positive samples classified incorrectly as negative. They both range between (0.0 and 1.0). Ideally both metrics should be 1.0, indicating that no negative samples were classified incorrectly as positive (high precision) and no positive samples incorrectly classified as negative (high recall).

The harmonic mean between precision and recall is entitled the F-measure (or F_1 score). This is a common metric employed in binary detection (Laroca et al., 2019; Powers, 2015) and can be calculated as follows:

$$F_1 = \frac{2}{P^{-1} + R^{-1}} = 2 \cdot \frac{P \cdot R}{P + R}. \quad (2.3)$$

F-measure is useful for obtaining a trade-off between precision and recall, considering that both of them have the same relevance for the systems performance. In detection tasks it can be the difference between wrongly detecting one more instance or missing an instance of an object.

It is important to point out that F-measure is not a recommended metric to assess multi-class detection performance (Powers, 2015). Therefore for those type of situations, Average Precision is a common metric reported in many famous detection works (Ren et al., 2017; Redmon et al., 2016). The Average Precision is a simplification of the precision/recall curve (Everingham et al., 2010), and can be calculated by taking the corresponding precision of 11 intervals of the recall uniformly distributed $[0.0, 0.1, 0.2, \dots, 1.0]$, formulated as

$$AP = \frac{1}{11} \sum_{r \in [0.0, 0.1, \dots, 1]} p_{interp}(r) \quad (2.4)$$

The precision at each recall level r is interpolated to avoid bumps that may be present in the PR curve. The interpolation maintains the maximum precision measured for any recall level $\tilde{r}$ that is bigger than r , maintaining monotonicity, as defined in Equation 2.5.

$$p_{interp}(r) = \max_{\tilde{r}: \tilde{r} \geq r} p(\tilde{r}) \quad (2.5)$$

As we are dealing with 10 classes (10 distinct digits), the mAP can be calculated, averaging the AP of every single class i , as follows:

$$mAP = \frac{1}{c} \sum_{i=1}^c AP_i = \frac{1}{10} \sum_{i=1}^{10} AP_i. \quad (2.6)$$

2.4 ADMR METRICS

Although mAP is interesting for measuring the quality of the detection methods, regarding pixel-wise accuracy, AMR systems focus on the detected reading sequence, and not on obtaining precise coordinates of the dials/meters. To best evaluate the performance of an ADMR system, three metrics are proposed: (i) Meter Recognition Rate, (i) Dial Recognition Rate, and (iii) Mean Absolute Error. The same metrics can also be directly applied to regular AMR systems, since the first two were previously adopted by Laroca et al. (2019) for digit-based meters.

Our main focus is to correctly recognize the final reading. The *Meter Recognition Rate*(MRR) evaluates the recognition of the final reading. The reading is a sequence of digits, and to determine if it was correctly recognized, we compare the predicted sequence ($pred_m$) and the ground-truth sequence (gt_m), for each of the N meters. If they are equal, the meter was correctly recognized (see Equation 2.7).

$$MR_{rate} = \frac{1}{N} \sum_{m=1}^N match(pred_m, gt_m) \quad (2.7)$$

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

For the *Dial Recognition Rate*(DRR), we employ the Levenshtein distance (also known as edit distance), a common measurement for computing distance between two sequences of characters. The Levenshtein distance measures the minimum number of edits (addition, removal or replacement of characters) required to transform one sequence in the other (Navarro, 2001). Levenshtein distance is suitable for our evaluation since it can handle small sequence errors in sequences that other metrics would treat as a big error. For instance, if we have a ground-truth sequence $a = "1234"$ and a prediction sequence $b = "234"$, a per-character evaluation metric would consider the error equal to 4, while Levenshtein distance is equal to 1, as the difference between them is a single missing pointer. The Levenshtein distance between the sequences a and b can be determined using the formulation

$$lev_{a,b}(i, j) = \begin{cases} \max(i, j), & \text{if } \min(i, j) = 0, \\ lev'_{a,b}(i, j) & \text{otherwise,} \end{cases} \quad (2.8)$$

where

$$lev'_{a,b}(i, j) = \min \begin{cases} lev_{a,b}(i-1, j) + 1 \\ lev_{a,b}(i, j-1) + 1 \\ lev_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)}. \end{cases}$$

The Levenshtein distance between the prediction and the ground truth is computed and then divided by the longest sequence size between them. This gives us the error. Subtracting the error from 1 gives us the dial recognition rate for each meter. Finally, the mean of all dial recognition rates yields the final dial recognition error, as defined in Equation 2.9.

$$DR_{rate} = \frac{1}{N} \sum_{m=1}^N 1 - \frac{lev(pred_m, gt_m)(|pred_m|, |gt_m|)}{\max(|pred_m|, |gt_m|)} \quad (2.9)$$

We emphasize that the sequence of digits that composes the meter reading is, in fact, a number (integer). Correctly predicting the last digit in the sequence is not as important as correctly predicting the first one (i.e., the most significant digit). In order to differentiate and penalize errors in the most significant digits, the *Mean Absolute Error* (MAE) is simple yet effective. After converting the sequences to integers ($pred_m$ and gt_m become the integers p_m and g_m , respectively), the mean absolute error of all N meters can be obtained using:

$$MA_{error} = \frac{1}{N} \sum_{m=1}^N |p_m - g_m|. \quad (2.10)$$

2.5 RESIDUAL NETWORKS

Classification networks can be used to correctly recognize the pointed values of a dial. One of the most recent breakthroughs in deep classification networks is ResNet (He et al., 2016). In contrast to regular networks that focus on learning a distribution $D(x)$ from an input x , the residual network's (ResNets) layers also focus on the residual part ($D(x) - x$). The residual learning process introduces lower level features directly to higher abstraction layers (through the use of shortcuts), preserving information and speeding up the learning process (see Figure 2.4). The introduction of the residual blocks enabled deeper networks architectures while having fewer parameters than shallower networks, such as VGG19 (Simonyan and Zisserman, 2015). ResNet also performs better and converges faster.

For the regular ResNet (He et al., 2016), the total number of parameters are

$$1 \times [256 \times 64 \times (1 \times 1) + 64 \times 64 \times (3 \times 3) + 64 \times 256 \times (1 \times 1)] \approx 70k \quad (2.11)$$

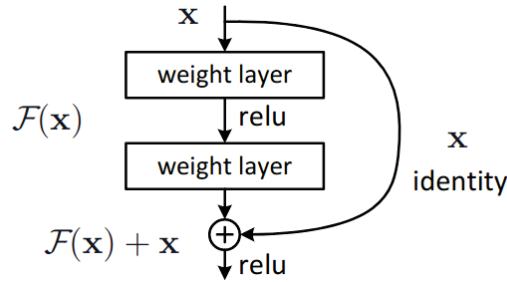


Figure 2.4: ResNet's residual block. The shortcut path (on the right) introduces lower level features directly to higher abstraction layers, speeding up the learning process and preserving information. Extracted from He et al. (2016).

Xception (Chollet, 2017) is inspired by the Inception models (e.g. Inception V3 (Szegedy et al., 2016) and explores the similarity between the Inception modules and Depthwise Separable Convolutions. A Depth-wise Separable Convolution (DSC) is a modified convolution developed in order to reduce the number of parameters required to perform a convolution. It consists in apply first a depthwise convolution (a channel-wise $n \times n$ spatial convolution), and then a pointwise convolution (1×1) convolution. The main objective is to avoid performing convolution across all channels, thus reducing the number of connections and calculations necessary.

The network is constructed by replacing regular Inception modules with Modified DSC. The only difference between the regular DSC and a modified DSC is the order of the operations, as show in Figure 2.5. The authors decided to invert the order of the pointwise and depthwise operations to be more similar to the regular Inception modules. Xception network trains almost

as fast and achieves better accuracy results in comparison with Inception V3, using a lower number of parameters.

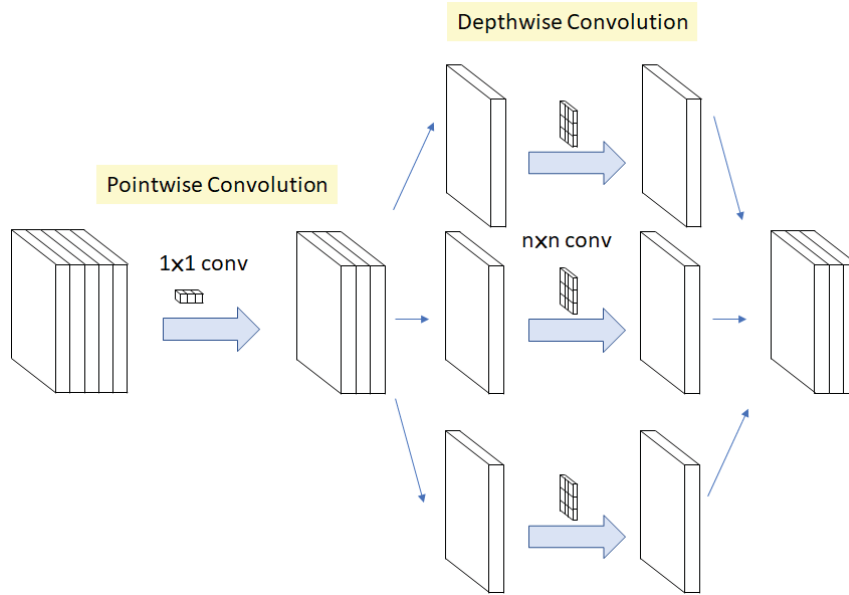


Figure 2.5: The Modified Depthwise Separable Convolution consists in first applying a pointwise convolution followed by a depthwise (channel-wise convolution). Extracted from Towards Data Science ¹.

2.6 YOLO

Redmon et al. (2016) proposed YOLO (You Only Look Once), an object detector that focuses on a speed/accuracy trade-off by dividing the input image into regions and predicting bounding boxes and probabilities for each region. The image is initially divided in an $S \times S$ grid (see Figure 2.6). Each grid cell is responsible for predicting B bounding boxes, containing their position (x, y) , size $(width, height)$ and the confidence score. For each cell and existing class, a class probability is also calculated. The final detection is performed using NMS, keeping the object's bounding boxes with higher confidence, removing intersecting ones ($IoU > 0.5$).

YOLOv2 (Redmon and Farhadi, 2017), an improved version of YOLO, adopts a series of concepts (e.g., anchor boxes, batch normalization) from existing works with novel concepts to improve YOLO's accuracy while making it faster (Liu et al., 2019a). Similarly, Redmon and Farhadi (2018) introduced YOLOv3 (the latest version of YOLO) which uses various tricks to improve training and increase performance, such as residual blocks, shortcut connections (or skip connections), and upsampling.

YOLOv4 (Bochkovskiy et al., 2020) introduced a novel backbone, along with a spatial pyramid pooling additional module to replace the Feature Pyramid Networks (FPN). The speed and accuracy performance were improved in comparison with YOLOv3. YOLO-based models have been successfully applied in several research areas (Liu et al., 2019a).

¹<https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568>

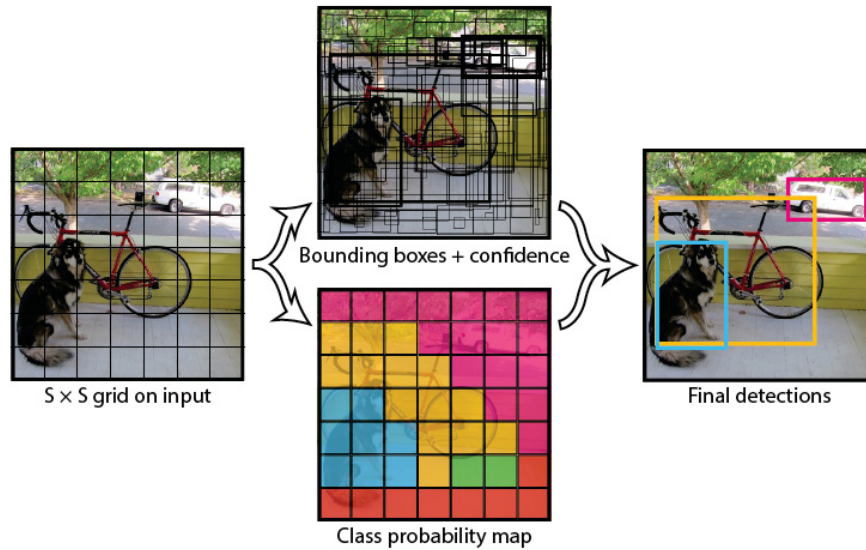


Figure 2.6: Illustration of the $S \times S$ grid. For each grid cell, B bounding boxes are generated, with their respective confidence. Besides, the class probability map illustrates the predominant class for each cell. Extracted from Redmon et al. (2016).

2.7 SEGMENTATION-FREE NETWORKS

We evaluated several segmentation-free models well known in the literature. Most of these models are commonly employed for scene text recognition — a task which main goal is to detect text regions in the complex background and label them with bounding boxes. Some of them are multi-task networks, i.e., those proposed or adapted by Gómez et al. (2018), Laroca et al. (2019) (Multi-task), and Calefati et al. (2019), while others are recurrent networks, more specifically Long Short-Term Memory (LSTM)-based networks, commonly employed for scene text recognition: CRNN (Shi et al., 2017), GRCNN (Wang and Hu, 2017), R2AM (Lee and Osindero, 2016), RARE (Shi et al., 2016), Rosetta (Borisjuk et al., 2018), STAR-Net (Liu et al., 2016a), TRBA (Baek et al., 2019), and ViTSTR (Atienza, 2021). In our work, LSTM-based methods are interesting to be evaluated since they provide a recurrent analysis of the images and sequences (they take into account information of the previously predicted value to predict the next value). Therefore, they may assist the ADMR to avoid prediction errors or correct them in case of misalignments (see Section 6.4 for more information on this type of error).

3 RELATED WORK

We review below several works that dealt with AMR and ADMR. For organization purposes this chapter is divided in sections according to the type of meters: single dial meters (also, pointer-type meters or gauges), multi-dial meters (in our case, energy meters) and digit-based meters (cyclometers and electronic for energy, gas and water consumption reading). Afterward, a review of the available datasets for AMR evaluation. Finally, we present an overview of the approaches.

3.1 SINGLE DIAL METERS

Zhao et al. (2005b) used a Radial Basis Function Neural Network (RBFNN) for component classification (dial, marks and needle) and Least Squares (LS) to determine the parameters of a function that can estimate the needle position and yield a correct measurement. In tests performed on 6 controlled images, the RBFNN surpassed a Multilayer Perceptron (MLP).

Jiale et al. (2011) introduced Improved Hough Line Transform (HLT), a method faster than regular HLT, by limiting the angle of line search. The paper uses morphological operations, Canny, adaptive thresholding and HT to yield a reading. It only tested on 4 images and there was no Region of Interest (RoI) detection (to detect the meter), only pointer detection.

Zheng et al. (2017) used a simple approach, including: binarization and HLT (for tilt correction) and recognition. The author evaluated the approach on only 3 controlled images of different types of pointers (2 for temperature and 1 for pressure).

In contrast to this simplistic binarization approach, Zhang et al. (2016) proposed a local binarization entitled Localized Pixel Inhomogeneity Factor (LPIF), creating more contrast and making the method more robust to the presence of noise. The proposed pipeline is a common approach in the single dial (gauge) recognition literature and is shown in Figure 3.1. Hough Circle Transform (HCT) detects the gauge and HLT recognizes the pointer. The final reading is obtained by the relation between the pointer angle and gauge scale. The dataset was challenging, but contained only 7 images which are not publicly available.

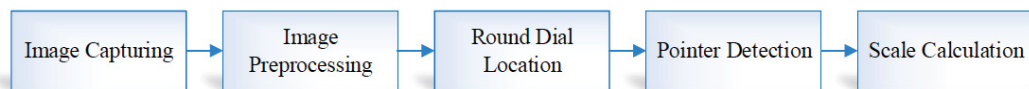


Figure 3.1: A common single dial (gauge) recognition pipeline. Extracted from Zhang et al. (2016).

Gao et al. (2018) used improved ORB feature matching to detect the pointers. The proposed method, surpassed Speeded Up Robust Features (SURF). After binarization, the final reading was performed using HLT. The images were strictly controlled and using a Pan-tilt-zoom (PTZ) camera (a camera that can move according to received commands). The number of images for testing was not available, as they were obtained and recognized on-the-fly.

Mai et al. (2018) used Scale-Invariant Feature Transform (SIFT) to match a predefined template to a gauge sample image to find the gauge, using k-Nearest Neighbors (kNN) and to reduce the matching error RANSAC (RANDOM SAMPLE CONSENSUS) is employed. At the end, a one-dimensional curve in the meter image is used, and the corresponding gray values of the pixels are used to find the pointer. The main disadvantages of the handcrafted method are: small dataset, low robustness to poor quality, noisy and uneven illumination.

Bao et al. (2019) used extrinsic parameters and inverse perspective mapping to correct perspective of the meter and improve final reading accuracy. First, it detects the marks on the meter scale to obtain the extrinsic parameters of the meter plane. Then, the extrinsic parameters can be used to correct the perspective of the meter using a transformation matrix. The final image is rectified and easier to recognize the pointed value. The authors also tested the method for multi-dial purposes (actually, only doubles dials). The main issues lie upon the fact that only a few images (less than a hundred) were evaluated, and the scenario was strictly controlled (lightwise).

Huang et al. (2019) applied deep learning for detection of the gauges using Region-based Fully Convolutional Network (R-FCN). After detection, an adaptive threshold converts the gauge to a binary image. A probability-circle method is used to segment the pointer and skeletonization and the HLT is used to find the pointer direction. Although the dataset is large (800 images), the main issue is that there is no solution for tilted and rotated gauge images. In addition, the binarization would not work in noisy and contrasting images.

Zhu et al. (2019) used Particle Swarm Optimization (PSO) to fit ellipses using arc segments. PSO performed better than HCT and Partial LS (see Figure 3.2). The number of tested images was not available as the accuracy in the final reading, only in the detection. No metrics regarding how the accuracy was calculated were provided neither.

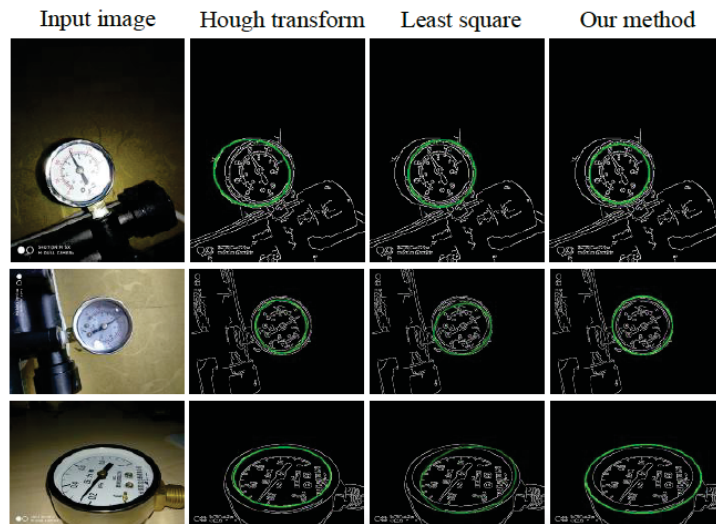


Figure 3.2: Comparison between HCT, LS and PSO. Adapted from Zhu et al. (2019).

Ren et al. (2019) used template matching, followed by an arc recognition strategy to detect and determine the gauge limits. After that, using a rotating window scanning around the central gauge point, it uses a score based in the pixel intensity to find the pointer. A homomorphic filter was applied to improve illumination issues caused by obtaining images on outdoor environments. No experiments were performed to evaluate the final reading, only qualitative experiments to verify if the pointer was correctly recognized.

Lauridsen et al. (2019) introduced a new approach and dataset for gauge pressure automatic reading. The methods used include gauge detection using LS circle fitting, binarization, scale marks recognition was made using features of binary objects and K-means, and finally the pointer angle was estimated using Principal Component Analysis (PCA). The final reading was not evaluated, only the pointer angle was used for testing. Even though the images were controlled and high resolution, the method was not robust to small variations, resulting in recognition errors or no recognition at all of some frames.

Liu et al. (2019b) evaluated Faster R-CNN, YOLO and Single Shot MultiBox Detector (SSD) for gauge detection. As expected, Faster R-CNN surpassed the other approaches. Besides the detection stage being deep learning-based, the rest of the reading pipeline is based on: HCT, transformation matrix (for perspective correction, see Figure 3.3), binarization and HLT for pointer recognition. The dataset, containing 9,218 high definition indoor images was not made available. The main difference of this approach was the use of active image acquirement, using a robot with a camera. In this sense, it was possible to obtain and combine several images, removing specular reflections and centered images (using a PTZ camera).

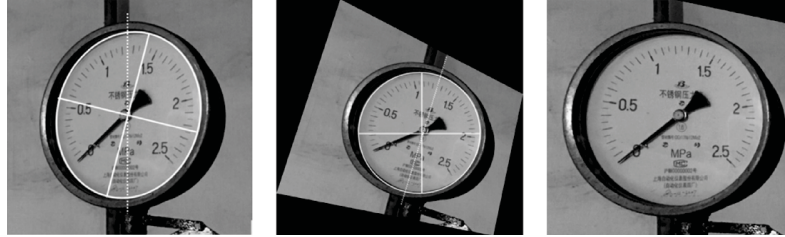


Figure 3.3: Perspective correction achieved using a transformation matrix. Extracted from Liu et al. (2019b).

Mask R-CNN was proposed for pointer recognition (Fang et al., 2019; He et al., 2019). Fang et al. (2019) used the network to find reference key points and the pointer in a gauge scale marks while He et al. (2019) focused on segmenting the meter dial and pointer.

Fang et al. (2019) modified Mask R-CNN to detect pixel key-points instead of masks, changing the loss function to accommodate the change. The key-points are basically scale marks to indicate the current measurement and the pointer location (4 key-points form a line). To perform the measurement, the mark scales are fitted into a circle using the LS method. The line that goes through the pointer's key point and the circle perimeter are intersected and the corresponding value is yield as the correct measure (as seen in Figure 3.4). For comparison, a Hough-based method (circle and lines) was tested, but the deep learning approach obtained a smaller absolute error. The dataset images were controlled (with small variations). Only one type of meter was evaluated and the meter was cropped to almost the image size, easing the recognition process.

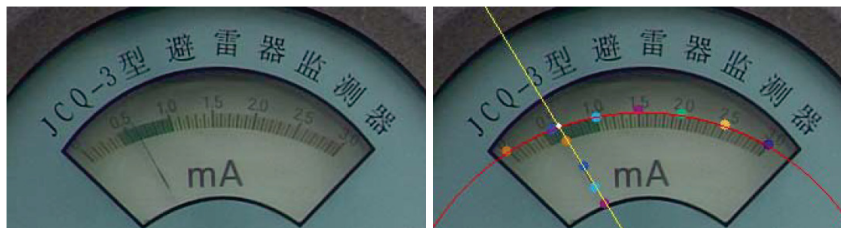


Figure 3.4: Illustration of the annotated points: one point for each scale mark, and five points for the pointer. Extracted from Fang et al. (2019).

In contrast, He et al. (2019) replaced the original RoIAlign layer of Mask R-CNN for PrRoIPooling to improve segmentation performance. The network detects a mask of the gauge, and a perspective transformation proposed by the authors based in the irregular ellipse detected helps to correct the pointer's perspective for final recognition. The pointer direction is calculated using PCA, and the final reading takes into account gauge information (e.g., scale and instrument position). The complete pipeline is displayed in Figure 3.5. Following this work, additional evaluation of the proposed methods and comparison with other works was performed (Zuo et al., 2020).

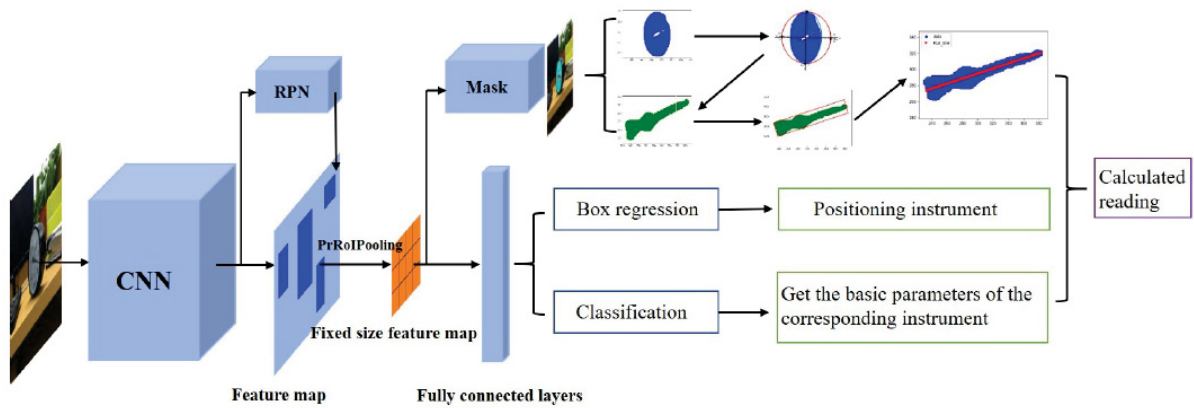


Figure 3.5: Pipeline for automatic gauge reading. The dial and pointer segmentation were performed with Mask R-CNN. After perspective correction, the pointed value is recognized using PCA and the gauge recognized parameters. Extracted from He et al. (2019).

Howells et al. (2021) recently introduced a real-time approach for gauge reading on mobile devices. They leveraged synthetic 3D-rendered gauge images to train two CNN detectors, one for gauge detection and one for gauge reading. The proposed system corrects the gauge's perspective and identifies the scale marks to correctly predict the reading. Although their system outperformed another method for gauge reading, it was evaluated exclusively on images with the gauge well centered and occupying a good 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.

3.2 MULTI-DIAL METERS

There are not many works that explored multi-dial meters, most of them focused on energy dial meters. For instance, Ocampo-Vega et al. (2013) used SIFT for detection of dials in energy meters. In sequence, morphological operations and trigonometric equations yield a final reading of energy dial meters (see Figure 3.6). Their approach was evaluated in 141 images in a once publicly available dataset (the page to obtain the dataset is not available anymore). Although the authors stated that the images were taken in real conditions, the quality of the images appeared to be controlled, with no great variations in light, angle and presence of dirt. The evaluation protocol and metrics were also not clearly described.

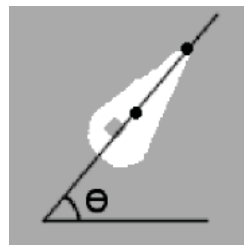


Figure 3.6: To perform the dial reading, the two points (centroid and the furthest point from it) were chosen, and a line defined by them is used for the angle calculation (the angle between the line and the x-axis). Extracted from Ocampo-Vega et al. (2013).

Souare et al. (2014) used a Laplacian filter to detect borders, affine transformation to perform registration (convert different images to the same coordinate system), binarization, morphological operations and angle calculation to monitor the energy usage in a dial meter. The

authors did not perform evaluation of the methods, nor provide the dataset, the input images were already cropped to the RoI and their conditions were strictly controlled.

Tang et al. (2015) focused in the constant monitoring of energy use using a smartphone connected to the Internet. The images were strictly controlled and the methods used to perform the recognition were trivial: binarization, morphological operations, and the pointer angle was determined using a trigonometric function. This method is not invariant to rotation, noise or other issues. The images were taken indoors every 15 minutes with a fixed camera (see Figure 3.7), and a remote system processed the data to generate estimates of energy consumption.



Figure 3.7: Images were captured every 15 minutes and send wirelessly to a remote server. Extracted from Tang et al. (2015).

3.3 DIGIT-BASED METERS

The main challenges for digit-based meters (cyclometers and digital meters) are the size of the digits and the possibility of Optical Character Recognition (OCR) systems confusing the meter digits with other text information printed on the meter.

Zhao et al. (2005a) was one of the first works to propose computer vision techniques for AMR. The techniques employed at the time, considered solely controlled images and handcrafted methods, such as binarization, shape matching methods and histogram for digit localization. The RoI location is hard-coded. Essentially, none of the conditions expected in a real world situation.

Nodari and Gallo (2011) introduced MultiNOD for analog gas meter reading. MultiNOD is a tree of MLPs, combined in such a way that each MLP produces a soft segmentation map, and several multi-scale maps combined produce a final RoI segmentation and subsequently the digits segmentation. The dataset was made publicly available and contains 153 images for RoI detection, 94 images for digit segmentation. The 332 digits correctly segmented yield an accuracy of only 58.73%. The final system only achieves a recognition rate of 30%, using 9 MLPs.

Vanetti et al. (2012) proceeded with the research of Nodari and Gallo (2011), enhancing the MultiNOD approach with the introduction of Fourier analysis to find the digits sequence and Support Vector Machine (SVM) to the final classification of digits. The MultiNOD is only used to find the RoI, segmenting the sequence of digits from the rest of the image. The Fourier analysis was used to detect individual digits, as it generates less non-digit artifacts than the MultiNOD. Finally, SVM is applied with a radial basis function kernel to recognize each detected digit. A

representation of each step's result is provided in Figure 3.8. The authors did not compare the new approach with the previous work. However, the experiments were made for each individual component, evaluating the impact in each stage. The final experiment – evaluating the whole system – was carried out in a novel dataset, obtaining 87% overall accuracy.

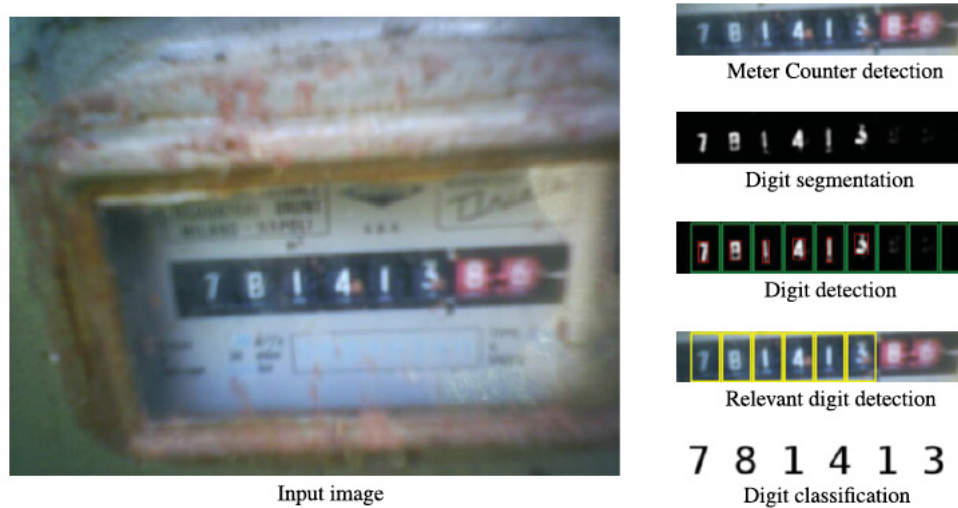


Figure 3.8: Illustration of each step of the meter recognition process. Extracted from Vanetti et al. (2012).

Gallo et al. (2015) dealt with gas meter reading, using cyclometer displays. They evaluated their approach in over 160k images. The method consists of using an MLP to locate the RoI. Digit detection is performed using Maximally Stable Extremal Regions (MSER) (a salient region detector) and the digits recognition with the use of Histogram of Oriented Gradients (HOG) for feature extraction and SVM for classification. The final reading error is 67.5%. The main reason for the poor performance lies upon the fact that MSER is not robust to poor illumination conditions and the recognition methods can make error due to blur, perspective and variations in illumination.

Gómez et al. (2018) used a segmentation-free approach for gas analog meter reading. A shallow CNN was employed (see Figure 3.9), with 5 distinct SoftMax layers that output the final prediction (numbers from 0 to 9 or none). The authors stated that the images present poor quality, however they do not provide samples to attest it and the dataset of over 200k images is closed. The segmentation approaches used for baseline is Haar cascade detectors. The recognition rate with the ground-truth bounding boxes achieve 95% recognition rate. The whole pipeline only achieves 54% recognition rate. It is senseless to use Haar cascade for comparison, when there were other methods such as YOLO and Faster R-CNN already available at the time. The segmentation-free approach achieves 94% recognition accuracy.

Laroca et al. (2019) achieved 100% F-measure on the digit segmentation task using Fast-YOLO (a shallower YOLO), in publicly available datasets, which indicates that detection (in this case, called segmentation) is not a big challenge anymore.

Tsai et al. (2019) used SSD (Liu et al., 2016b) for RoI detection on cyclometer energy meters, but no digit segmentation or recognition was performed. The authors stated that there was no other work in the literature that uses a state-of-the-art detection model to detect the digits RoI, which is not accurate, since Laroca et al. (2019) uses YOLO for RoI detection. There was also no validation data, the training stops when the validation loss increases. The authors also used the image errors on testing to retrain the network in order to achieve 100% accuracy.

Laroca et. al. (Laroca et al., 2019) designed a two-stage approach for AMR. The Fast-YOLOv2 model (Redmon and Farhadi, 2017) was employed for counter detection and three

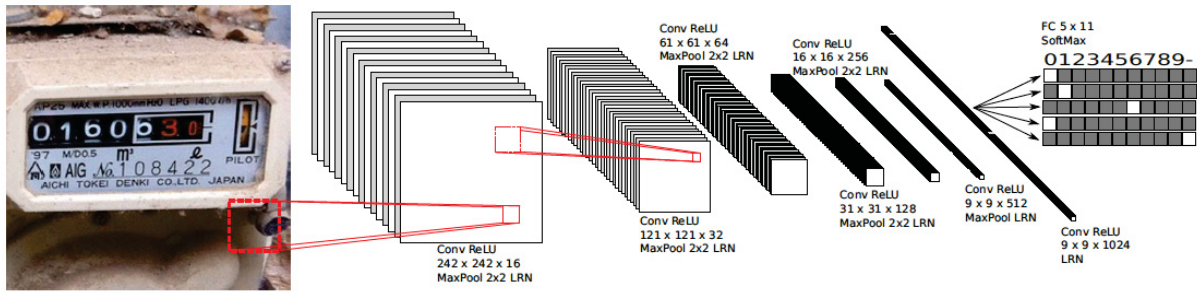


Figure 3.9: The shallow segmentation-free network. Extracted from Gómez et al. (2018).

CNN-based models were evaluated for the counter recognition task: CR-NET (Silva and Jung, 2017a), a multitask approach and CRNN (Shi et al., 2017). The authors considerably improved their recognition results when balancing the training set in terms of digit classes through data augmentation techniques. In their experiments, impressive results were achieved in digit-based meters, however, the images they used were acquired in a warehouse of the service company by one of the authors, i.e., the images were captured under controlled conditions.

Kanagarathinam and Sekar (2019) supplied the dataset for electronic energy meter reading. The method employed for the reading is handcrafted and the dataset is small and controlled, with high resolution up-close images and the RoI is manually cropped. Because of that, the segmentation task is only related to the digit's segmentation, not the meter's. The length of the meters range from 4 to 8 digits. The variations of the dataset are: tilt, blur and period of the day. But the changes are not significant as we can see in Figure 3.10. The method applied to segment is the MSER algorithm, a common blob detector, that works well in this case due to the uniform illumination of the images. For the digit recognition stage, the common OCR algorithm entitled Tesseract (Smith, 2007) was employed.



Figure 3.10: Electronic meter images. Extracted from the proposed dataset Kanagarathinam and Sekar (2019).

3.4 DATASETS

Most of the referred works do not provide a public dataset to enable a fair comparison of results. There are only a few publicly available meter datasets (Nodari and Gallo, 2011; Laroca et al., 2019; Kanagarathinam and Sekar, 2019; Lauridsen et al., 2019). Other works also proposed datasets (Zuo et al., 2020; Ocampo-Vega et al., 2013; Howells et al., 2021), but no functioning access links were provided and contact could not be established with the authors. As far as we

know, our UFPR-ADMR-v1 (Salomon et al., 2020) dataset is the only currently available dataset containing images of energy or multi-dial meters.

3.5 OVERVIEW

We compiled the approaches of the last sections for better visualization in Table 3.1. The aspects considered are: the problem being solved, the method employed for detection, the method applied to recognition, the usage of deep learning approaches (DL), number of images present in the evaluation data, and if the images were in a constrained or unconstrained environment. Note that no performance metrics are reported. Due to the inability of fair comparison between the works, accuracy/error results could be deceiving regarding the methods' performance.

Table 3.1: Summary of the reviewed works.

Paper	Problem	Det. Method	Recog. Method	DL	Imgs.	Env.
Single Dial Meter						
Zhao et al. (2005b)	meter calibration	RBFNN and MLP		Yes	6	C
Jiale et al. (2011)	multimeters	-	Bin + Improved HLT	No	4	C
Zhang et al. (2016)	gauges	LPIF + HCT	HLT	No	7	U
Zheng et al. (2017)	gauge instruments	OTSU Bin.	HLT	No	3	C
Gao et al. (2018)	substation gauges	ORB algorithm	Hough Line	No	-	C
Mai et al. (2018)	substation gauges	SIFT	KD-Tree + RANSAC	No	21	C
Bao et al. (2019)	measure gauges	Extrinsic Values + Invers. Mapping, Rotat. Matrix + HLT		No	48	C
Fang et al. (2019)	substation gauges	Mask R-CNN	ResNet-18 + LS	Yes	1160	+/-
Huang et al. (2019)	gas stations	RFCN	Bin + Prob. Circle + Skelet. + HLT	Yes	800	+/-
Zhu et al. (2019)	gauges	PSO	Hough Line	No	-	C
Lauridsen et al. (2019)	gauges	LS circle fitting	Bin. features + K-Means + PCA	No	7*	C
Ren et al. (2019)	meter reading	Template + Arc Recog.	window function scanning	No	-	U
He et al. (2019)	gauges	Mask R-CNN	Mask R-CNN + Perspect. + PCA	Yes	1,430	+/-
Zuo et al. (2020)						
Liu et al. (2019b)	gauges	Faster R-CNN	Perspect. + Bin. + HCT + HLT	Yes	9,218	C
Howells et al. (2021)	gauges	CNN	CNN	Yes	11K	C
Multi-dial Meter						
Ocampo-Vega et al. (2013)	dial energy meter	SIFT	bin + ang. calculation	No	141	C
Souare et al. (2014)	dial energy meter	rectangular crop	Filters + Transformations	No	-	C
Tang et al. (2015)	dial energy meter	rectangular crop	bin + trigonometric function	No	500+	C
Cyclometer						
Zhao et al. (2005a)	energy meter	Central Region of Image	bin + histogram + patter similarity	No	-	-
Nodari and Gallo (2011)	gas meter	MultiNOD	Tesseract	Yes	247	U
Vanetti et al. (2012)	gas meter	MultiNOD + Fourier	SVM	Yes	100	U
Gallo et al. (2015)	gas meter	MLP	MSER + HOG + SVM	Yes	160K	U
Gómez et al. (2018)	gas meter	No detection	CNN	Yes	200K	C
Tsai et al. (2019)	energy meter	SSD	-	Yes	200	C
Cyclometer and Electronic Meter						
Laroca et al. (2019)	energy meter	Fast-YOLOv2	CRNET, Multitask and CRNN	Yes	2K	+/-
Electronic Meter						
Kanag. and Sekar (2019)	energy meter	MSER	Tesseract	No	169	C

*Videos.

4 THE UFPR-ADMR-V2 DATASET

We acquired the meter images from Copel, the energy company of the state of Paraná in Brazil, that serves more than 4 million consuming units per month (Copel, 2020; Laroca et al., 2019). The images of the meters were obtained at the consuming units by Copel employees using smartphone cameras. All images had already been resized and compressed for storage, resulting in images of 640×480 or 480×640 pixels (depending on the image orientation). In our previous published work (Salomon et al., 2020), we provided a dataset containing 2,000 annotated dial images entitled UFPR-ADMR-v1 dataset. To create the UFPR-ADMR-v2 dataset, we selected 5,000 images in which it was possible for a human to recognize the correct reading of the meter, as the images were acquired in uncontrolled environments and it would not be possible to label the correct reading in many cases (see Figure 4.6). Figure 4.1 shows some of the images selected for the dataset.



Figure 4.1: Examples of the images chosen for the dataset. Note that the customer meter identification is blurred for keeping subject privacy.

In each image, we manually labeled the position (x, y) of each corner of an irregular quadrilateral that contains all the dials. After the corners are annotated, we rectified the quadrilateral and proceeded to label each dial using ellipses. At the end, we convert the ellipses to the original (non-rectified) image, and record the minimum bounding rectangle around each ellipse. The ellipses are also saved as segmentation masks. The annotation process is illustrated in Figure 4.2.

All meters have 4 or 5 dials, so that 2,590 meters (51.8%) have 4 dials and 2,410 meters (48.2%) have 5, totaling 22,410 annotated dials. The labeling process took more than 350 hours to be completed, not including the time required to improve the labeling process, the relabeling of several images and the selection stage — in this stage we removed all the images

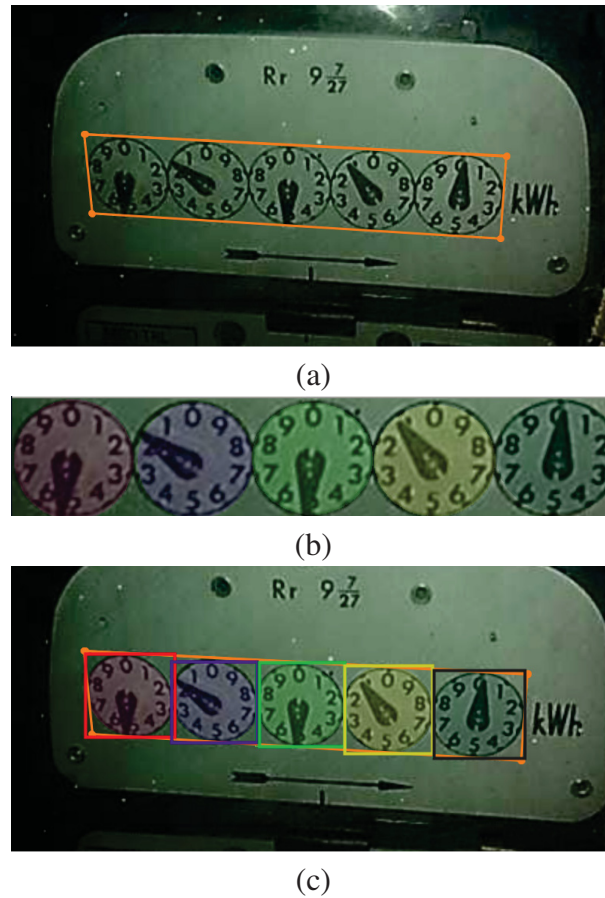


Figure 4.2: The labeling process with our custom made annotation tool consists of the following steps: (a) estimating the four corners of the irregular bounding rectangle; (b) rectifying the rectangle for annotation of ellipses; (c) generating the final bounding boxes in the original (non-rectified) image around each ellipse, together with the irregular bounding rectangle.

that did not contain dial meters or in which the dial reading could not be properly identified by a human. In addition to the irregular quadrilateral, the segmentation masks and bounding boxes, we also labeled the correct reading (sequence of integers ranging between 0 and 9) and the approximated pointed value (sequence of rational numbers with one decimal place precision, ranging between 0.0 and 9.9).

Information regarding the dimensions of the meters and dials in the dataset are available in Table 4.1. Note the great variability in the size of both meters and dials, for example, the smallest dial (20×29 pixels) is 10 times smaller than the largest one (201×206 pixels).

Table 4.1: Statistics about the size of meters and individual dials.

	Min (px) W $\times$ H	Max (px) W $\times$ H	Mean (px) W $\times$ H	Mean Area (px ²)
Meters	109×33	691×157	305×110	33,799
Dials	20×28	233×241	79×81	6,476

The values pointed on each dial have an almost uniform distribution of digits, having slightly more 0s on the first dial than other digits. The distribution of digits per dial is illustrated in Figure 4.3. The most prominent bar indicates that the most frequent digit in the first position is 0. Nevertheless, it should be noted that the distribution is not as unbalanced as datasets with

digit-based meter images, such as the UFPR-AMR dataset (Laroca et al., 2019), in which the number of 0s in the first position is equal to the quantity of 0s in all of the other positions combined, as we can see in Figure 4.4. This is probably due to the fact that dial meters stopped being manufactured and deployed decades ago, which implies that each dial may have completed many cycles since the installation and may be indicating any value.

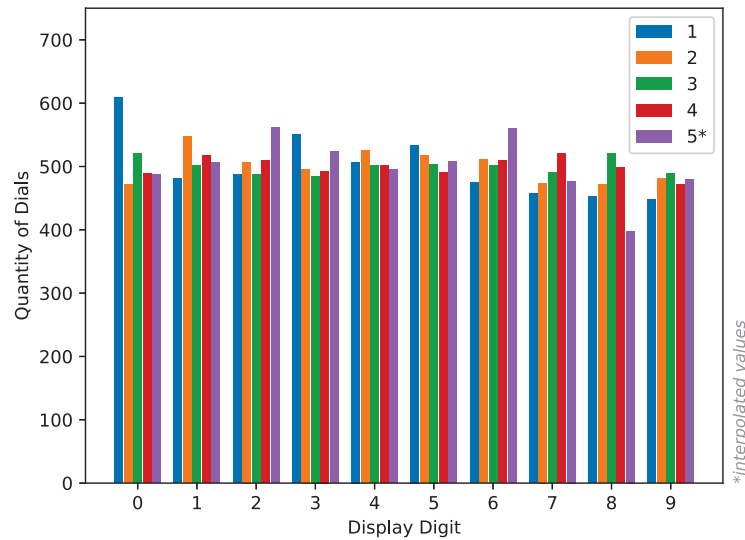


Figure 4.3: The distribution of digits according to the dial position on the meters from the UFPR-ADMR dataset. As 51.8% of the meters in the proposed dataset do not have a 5th dial, the values of the 5th dial quantities were interpolated for better visualization.

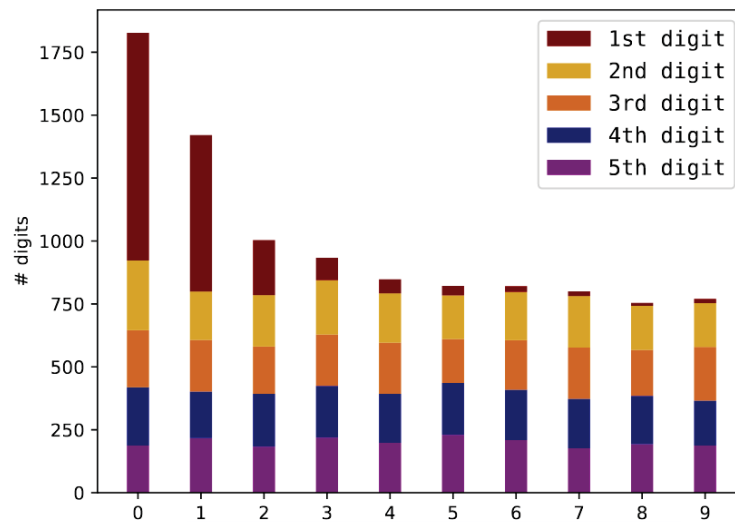


Figure 4.4: The distribution of digits according to the dial position on the meter, on the digit-based meters from UFPR-AMR dataset (Laroca et al., 2019).

Figure 4.5 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.

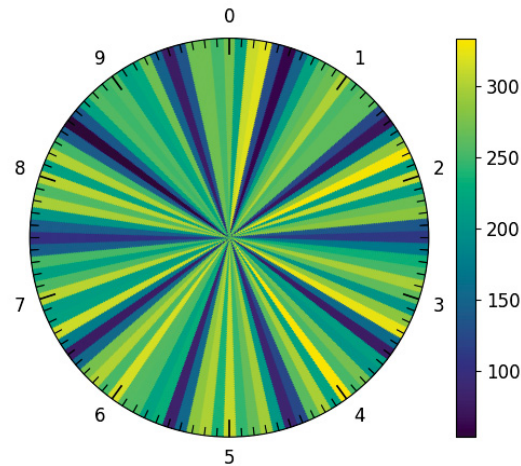


Figure 4.5: Distribution of the pointer positions in the UFPR-ADMR-v2 dataset.

4.1 CHALLENGES

The main challenge of the proposed dataset is the quality of the images. Low-end cameras, challenging environmental conditions and high compression are factors that degrade image quality. The challenging environmental conditions include: reflections, dirt, and broken glass, and low-quality acquisition may result in: noisy, blurred and low-contrast images. Figure 4.6 illustrates the main image-quality issues described above.

In addition to the aforementioned quality issues, there are several types of meter templates and each manufacturer has its own dial model (with variations on the scale marks spacing) and pointer design. This variations combined with the image capture angle make it difficult to determine the exact pointed value.

Another challenge arises from the presence of clockwise and counter-clockwise dials – for design purposes, each meter has alternating clock directions –, and the direction of the dials may differ depending on the meter model and manufacturer. Also, the dials can be positioned in a straight line or in a semi-circular manner (see Figure 4.6g).



Figure 4.6: Samples of the challenging scenarios present on the provided images. We selected for the UFPR-ADMR-v2 dataset 5,000 images in which it was possible for a human to recognize the correct reading of the meter. Note that most of the dials presented above are not readable, and were not selected for the dataset. We blurred the region containing the consumer unit number in each image due to privacy constraints. None of the images above were selected to our database, they are displayed solely for illustration purposes.

5 PROPOSED APPROACH

We evaluated three different approaches for ADMR: (i) object detection with recognition; (ii) object detection with regression; and (iii) segmentation-free recognition. The main motivation to consider object detection is the compelling results achieved in several previous works (Laroca et al., 2021a; Salomon et al., 2020; Laroca et al., 2019). The regression approach is also interesting for it increases the prediction granularity with rational numbers instead of integer predictions. Consequently, we can estimate whether the pointed value is close to the scale mark or not, enabling the development of correction methods. Lastly, we explored segmentation-free methods. Methods such as LSTM are commonly applied in the literature to retrieve sequences of characters(e.g. text, numbers), specially for scene-text recognition (Borisyuk et al., 2018; Liu et al., 2016a; Baek et al., 2019; Atienza, 2021). The pipeline for each approach is illustrated in Fig 5.1.

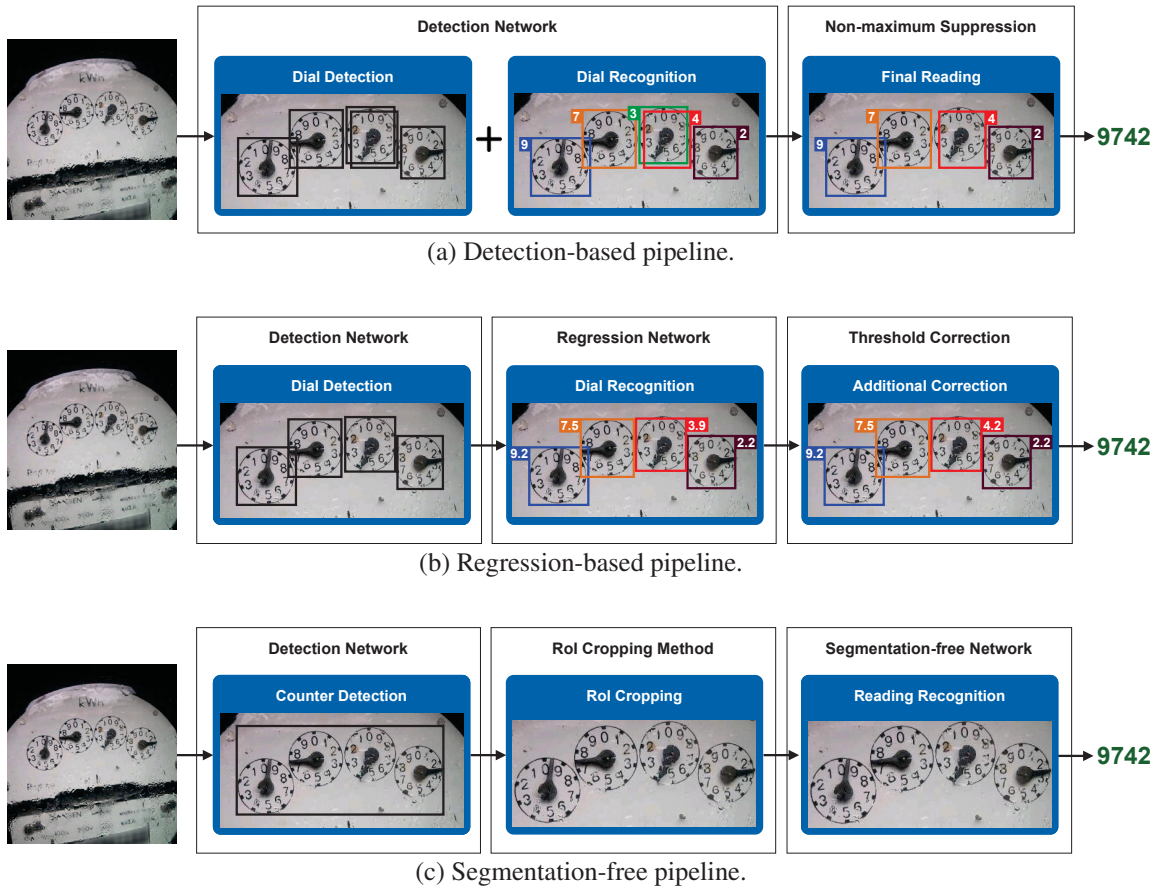


Figure 5.1: Proposed pipelines.

5.1 OBJECT DETECTION WITH RECOGNITION

As can be seen in Figure 5.1a, 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 (i.e., 0–9) using the bounding box of each dial as input. Afterward,

we apply a NMS algorithm to eliminate redundant detections (those with $\text{IoU} \geq 0.5$) since the same dial might be detected/recognized more than once by the network.

For this approach, we decided to evaluate YOLOv3 (Redmon and Farhadi, 2018) and YOLOv4 (Bochkovskiy et al., 2020) considering that they have been successfully employed for various computer vision tasks and have achieved promising reading results in digit-based meters (Liao et al., 2019; Laroca et al., 2021a).

Shallower detection models, such as those used by Laroca et. al. in (Laroca et al., 2019) (CR-NET) and (Laroca et al., 2021a) (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.

5.2 OBJECT DETECTION WITH REGRESSION

Similarly, regression approaches also leverage an object detector (Fast-YOLOv4-SmallObj (Laroca et al., 2021a)) to obtain the position of all dials from the input image (see Fig 5.1b). After the detection stage, fine-tuned ResNet (He et al., 2016) and Xception (Chollet, 2017) models are explored to predict a float number (ranging from 0.0 to 9.9) for each detected dial. ADMR can be addressed as a regression problem since we could introduce more precision to the pointer prediction by using rational values.

For training the 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. For the tangent function, we converted the labels values to angles 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):

$$\theta \mapsto (\sin(\theta), \cos(\theta)). \quad (5.1)$$

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 an input cropped dial, we need to convert those two cartesian coordinates to polar coordinates (in our case, to an angle θ).

And to convert the coordinates back from cartesian to polar we use the following formulation:

$$(y, x) \mapsto \arctan 2(y, x) \quad (5.2)$$

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

Afterward, we designed 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 Figure 5.1b. 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.

Figure 5.1b 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 (where N is a discrete label between 0 and 9), and the lower correction threshold for the next dial is 2.5. As $3.9 > N.75$ (considering only the fractional part) 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.

5.3 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 to detect/segment each of its gauges. Such models have been successfully explored in OCR applications (Calefati et al., 2019; Atienza, 2021).

For this approach, it is necessary to first detect the RoI (see Fig. 5.1c) 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 (Laroca et al., 2021a), called Fast-YOLOv4-SmallObj, as it was robust enough to reach 100% F-score ($\text{IoU} \geq 0.5$) in our experiments.

We then evaluated several well-known segmentation-free models in the literature. Some of them are multi-task networks, i.e., those proposed or adapted by Gómez et al. (2018), Laroca et al. (2019) (Multi-task), and Calefati et al. (2019), while others are recurrent networks, more specifically LSTM-based networks, commonly employed for scene text recognition: CRNN (Shi et al., 2017), GRCNN (Wang and Hu, 2017), R2AM (Lee and Osindero, 2016), RARE (Shi et al., 2016), Rosetta (Borisyuk et al., 2018), STAR-Net (Liu et al., 2016a), TRBA (Baek et al., 2019), and ViTSTR (Atienza, 2021). For all of the aforementioned methods, the maximum number of predicted number sequences is set to 5.

5.4 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. To this end, we used Albumentations (Buslaev et al., 2020), which is a well-known Python library for image augmentation.

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 4. However, contrary to what we expected, there was no significant improvement in the results achieved (further investigation is needed in this regard).

5.5 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° , as shown in Figure 5.2. Lastly, we feed the corrected image to the network and consider these predictions as the final reading.

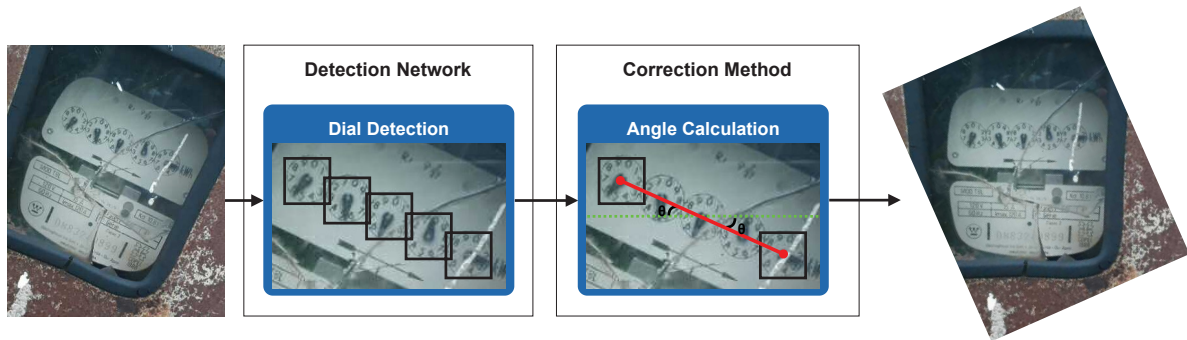


Figure 5.2: Illustration of the angle correction method. The angle (θ) between the line that passes through the middle of the first and last detected dial and the x -axis is computed. Afterwards, the correction method rotates the image $-\theta$ degrees in order to make the dials parallel to the x -axis.

6 EXPERIMENTAL RESULTS

This chapter is organized as follows. In the next section, we introduce the training and evaluation metrics and hardware used for training. Afterwards, we present the results for dial detection, meter recognition, dial recognition and Mean Absolute Error (MAE) for all evaluated approaches. In Section 6.4, we define and analyze the types of errors, their distribution, display results and provide solutions to improve performance and reduce errors in a real-world system. Lastly, we give a brief overview of the direction of future works showing some synthetically generated images using Generative Adversarial Networks (GANs).

6.1 TRAINING ENVIRONMENT AND EVALUATION METRICS

Most of our experiments were carried out on an AMD Ryzen Threadripper 1920X 3.5GHz 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.

The YOLO-based models were trained using the Darknet framework¹, the multi-task models were trained using Keras², and the segmentation-free models were trained using a fork of the open source repository of Clova AI Research³, which was used to record the 1st place of ICDAR2013 focused scene text and ICDAR2019 ArT, and 3rd place of ICDAR2017COCO-Text and ICDAR2019 ReCTS (task1).

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 (Ren et al., 2017; Redmon et al., 2016; Redmon and Farhadi, 2018).

We evaluated several metrics such as Root Mean Square Error (RMSE), Mean Squared Error (MSE), and MAE for the regression approaches. The metrics were used to assess the prediction error directly in the regression networks outputs (in the encoded labels, not in the final reading predicted value). The best results were achieved by using the MSE, which can be obtained by:

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad (6.1)$$

6.2 EVALUATION PROTOCOL

An evaluation protocol is necessary to enable fair comparison between different approaches. The dataset was randomly divided into three disjoint subsets: 3,000 images for training (60%), 1,000 images for validation (20%) and the remaining 1,000 images for testing (20%). This random division is fixed and used in every evaluation performed. Following recent works in which datasets were introduced (Laroca et al., 2019; Yang et al., 2019; Salomon et al., 2020), the subsets generated are explicitly available along with the UFPR-ADMR-v2 dataset.

The *mAP* will not be used for evaluation of the test set, as it is not important for us the pixel-wise quality of detection, but rather the sequence of pointed values. For assessing only

¹<https://github.com/AlexeyAB/darknet/>

²<https://keras.io/>

³<https://github.com/roatienza/deep-text-recognition-benchmark/>

detection performance, the precision, recall and F-score are employed. The previously discussed metrics (see Section 2.4): Dial Recognition Rate (DRR), MRR and MAE are used for assessing the system’s overall performance.

6.3 RESULTS

First, we investigate the performance of the models in the dial detection task. The results are listed in Table 6.1. For comparison, a common method proposed in the literature was evaluated: HCT (Zhang et al., 2016; Zheng et al., 2017). F-score was chosen as the evaluation metric, as it is often used to assess binary detection tasks. The metrics take into account the detection or not of the dials. If the $IoU > 0.5$ between the label and predicted dial area, a dial is considered detected. As expected, deep learning-based methods (i.e., YOLO) outperformed HCT achieving high F-score values. HCT does not handle well the large variations on lighting, contrast and perspective found in our dataset images.

Table 6.1: Dial detection results achieved on the UFPR-ADMR-v2 dataset.

Detection Model	Backbone	(%)		
		Prec.	Recall	F-score
Hough Circle Transform	-	48.70	51.15	49.89
Fast-YOLOv3	Darknet	99.94	100.0	99.97
YOLOv3	Darknet-53	100.0	100.0	100.0
Fast-YOLOv4	Darknet-53	100.0	100.0	100.0

The results obtained by all approaches and models evaluated are shown in Table 6.2. As can be seen, in general, the detection-based models (e.g., YOLOv4) outperformed both regression and multi-task 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 (Gonçalves et al., 2018; Laroca et al., 2021b). This is due to the fact that two-stage methods usually propagate the error from one stage 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 (Gómez et al., 2018; Laroca et al., 2019). A solution to this lack of data and reduction of bias is proposed in Section 6.5.

The best results were achieved by combining the best detection model (YOLOv4 (Bochkovski et al., 2020)) with the best regression model (Xception (Chollet, 2017)). This ensemble of distinct targeted networks is our strategy of correction, where Xception indicated when the corrections should be performed, reducing the errors in the right-most dials. Previously, we evaluated several combination of detection networks in order to indicate when the prediction should be corrected. However, the results were always worse than just using a single detection network without correction. The main reason is that the detection networks could not indicate where exactly the pointer was located (most of our errors occur near the dial’s ticks); thus, the detection network was more precise on that matter.

The best results achieved in this work are significantly better than those achieved in our previous work (Salomon et al., 2020). The reasons are threefold: (i) there are now more than twice as many images for training the deep networks, which highlights the importance of the proposed 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 (Salomon et al., 2020): neighbor values, rotated images, lighting, and mirrored predictions, as described in the following subsection. Some samples of the successfully predicted images are displayed in Figure 6.1.

Table 6.2: Reading results achieved on the UFPR-ADMR-v2 dataset.

Approach	Meter Recog.	Dial Recog.	MAE
Laroca et al. (2019) (Multi-task (Gonçalves et al., 2018))	15.30%	64.24%	24,181
Calefati et al. (2019)	21.00%	72.20%	20,078
Gómez et al. (2018)	23.50%	70.56%	24,074
Rosetta (Borisjuk et al., 2018)	57.50%	82.39%	4,968
R2AM (Lee and Osindero, 2016)	59.60%	83.72%	4,087
STAR-Net (Liu et al., 2016a)	60.00%	83.70%	9,454
GRCNN (Wang and Hu, 2017)	60.80%	84.14%	6,108
RARE (Shi et al., 2016)	61.80%	84.25%	5,696
CRNN (Shi et al., 2017)	64.20%	86.27%	3,669
TRBA (Baek et al., 2019)	66.80%	87.87%	2,404
Laroca et al. (2019) (CR-NET (Silva and Jung, 2017b))	71.40%	91.94%	1,680
ViTSTR-Base (Atienza, 2021)	76.20%	92.45%	2,841
Laroca et al. (2021a)	78.30%	94.07%	1,327
ResNet50 (He et al., 2016)	85.00%	96.38%	429
ResNet101 (He et al., 2016)	86.20%	96.52%	402
Xception (Chollet, 2017)	88.00%	96.96%	244
YOLOv3 (Redmon and Farhadi, 2018)	88.60%	97.00%	477
YOLOv4 (Bochkovskiy et al., 2020)	91.10%	97.67%	242
YOLOv4 (Bochkovskiy et al., 2020) (Rotated)	92.00%	98.00%	197
YOLOv4 (Bochkovskiy et al., 2020) (Rotated) + Xception (Chollet, 2017)	92.50%	98.11%	187

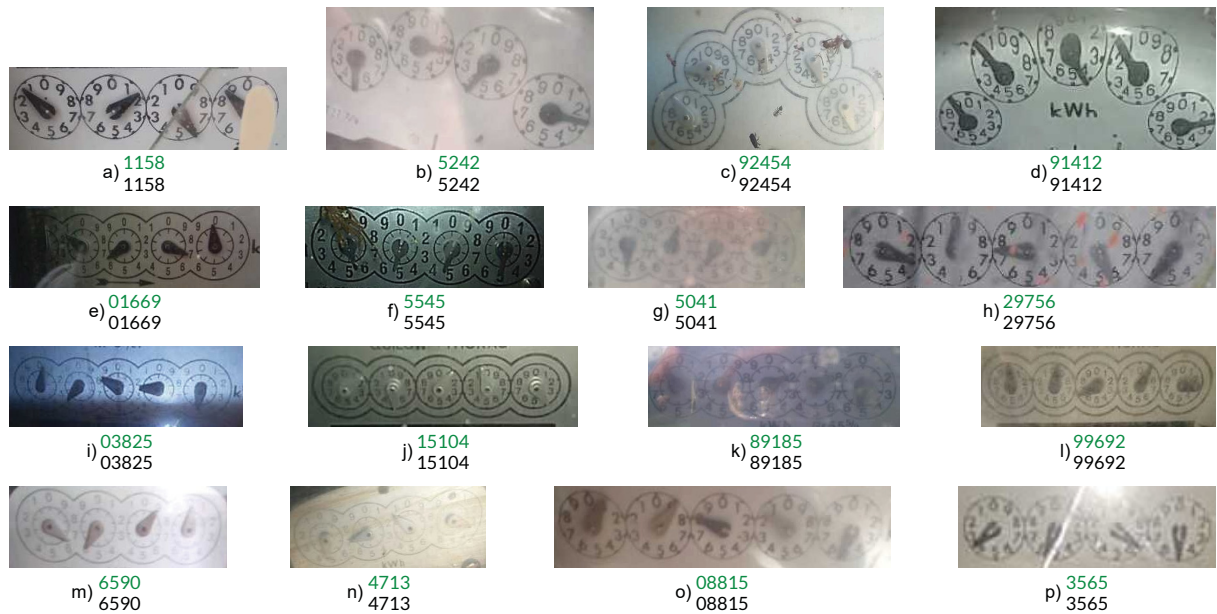


Figure 6.1: Prediction (top number) and ground-truth (bottom number) examples of correctly recognized meters (the images were cropped for better visualization).

6.4 ERROR ANALYSIS

To better assess the problems obtained by our approaches, it is important to analyze the sources of system errors. After meticulous inspection of the wrongly recognized meter images, we conclude that the most common errors were caused by:

- **Symmetry:** as there are clockwise and counterclockwise dials, sometimes when the digits are blurred, the method can not differentiate the direction and thus may output the mirrored value of the real prediction.
- **Neighbor value:** the most common error. Variables such as: angle, lighting, shadows and occlusion (when the pointer is in front of the dial mark) can hinder the reading of a dial. There were even disagreements between authors regarding the correct pointed value, in such situations (Salomon et al., 2020). Some dials are also **misaligned** in comparison with the next ones (this might be caused by mechanical failure, manufacturing defects, weathering or other causes).
- **Severe lighting conditions/Dirt:** shadows, glares, reflections and dirt may confuse the networks, specially 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. Dirt in the meter glass may interfere in the reading process causing this low contrast.
- **Rotation:** rotated images are harder to predict, since the pointed value is not on the usual position. The predictions may be assigned to neighbor digits that, would be in the current angle of the pointer if the image was not rotated.

To illustrate the sources of errors, some of them are displayed in Figure 6.2. The neighbor values issue was tackled by combining the regression with the detection approach. The regression approach uses the neighbor value to correctly yield a prediction, thus achieving a lower MAE even though the MRR and DRR are slightly lower than YOLOv3. This approach was only possible due to the new type of annotation provided in our new UFPR-ADMR-v2 dataset.

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) + Xception).

As aforementioned, 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 YOLOv4's recognition rate achieved by 0.9%.

As listed in Table 6.3, most errors occurred on the least significant dial (i.e., the rightmost one). Although there are multi-dial errors, as shown in Figure 6.2, these errors have an Absolute Error of 1. In fact, 85% of all errors have an Absolute Error of 1. We estimate that such an error would only cost approximately R\$ 0,83 to the regular customer⁴ (or to the company) and would probably be compensated in the next monthly reading. Moreover, it's hard even for a human to identify the correct reading, due to perspective, lighting, shadows, image quality. There were disagreements between authors regarding the correct label in many cases (Salomon et al., 2020), which makes this type of error really hard to avoid without a policy of rounding up or down the reading in such scenarios.

We also developed a simulation of a system which rejects errors higher than a predefined absolute value in KWh. This simulation uses our best results (achieved by YOLOv4 (R) +

⁴<https://www.copel.com/hpcweb/copel-distribuicao/taxas-tarifas/>

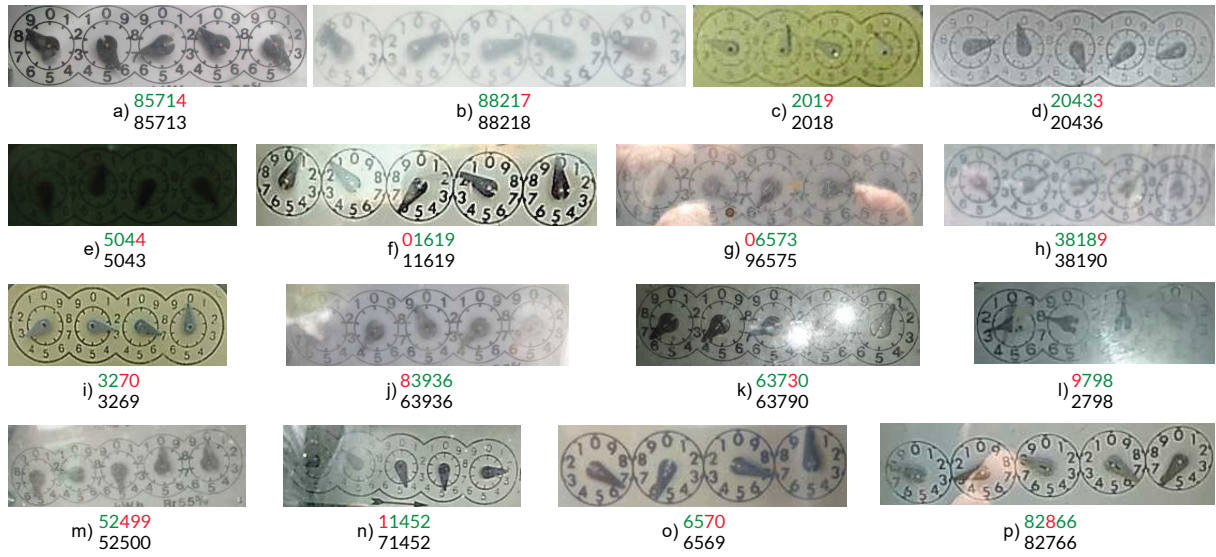


Figure 6.2: Prediction (top number) and ground-truth (bottom number) examples (the images were cropped for better visualization) of the errors of the best method (YOLOv4 (R) + Xception). The errors are marked in red and include: neighbor values (a,b,c,e,h,i,m,o); symmetry (d); misalignment (f,p); severe lighting conditions (g,j,k,n); and dirt(l).

Table 6.3: Distribution of errors by dial position.

Approach	Frequency (%)				
	1	2	3	4	5
YOLOv3 (Liao et al., 2019)	7.86	8.57	7.14	11.43	65.00
YOLOv4 (Bochkovskiy et al., 2020)	8.49	5.66	5.66	15.09	65.09
YOLOv4 (Bochkovskiy et al., 2020) (Rotated)	5.21	3.12	6.25	12.50	72.92
YOLOv4 (Bochkovskiy et al., 2020) (Rotated) + Xception (Chollet, 2017)	4.71	1.18	3.53	10.59	80.00

Xception) to estimate how the acceptance of errors improve the MRR. Thus, we can have a trade-off between tolerated error, its estimated cost and the required number of manual inspections. Note that by tolerating only 1kWh error, we can increase the MRR from 92.50% to 98.90% (at a cost of R\$0.83 per tolerated error). In this way, we consider that the YOLOv4 (R) + Xception model can be reliably employed on real-world applications.

Finally, it is worth noting that the service company could easily detect the most significant errors using simple heuristic rules or even with 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 should be double-checked by employees.

Table 6.4: The trade-off between accuracy and cost of error using YOLOv4 (Rotated) + Xception and the number of manual inspections required to perform all correct readings in the testing dataset (1,000 dials images).

Tolerated Error (kWh)	Meter Recognition Rate	Tolerated Cost(R\$)	# Manual Inspec. Req.
0	92.50%	0.00	75
1	98.90%	0.83	11
5	99.10%	4.16	9
10	99.30%	8.32	7
50	99.40%	41.58	6
100	99.50%	83.16	5

6.5 GAN GENERATED DATASET

We also employed a GAN entitled Pix2Pix (Isola et al., 2017) in order to generate more – and more balanced – images to perform data augmentation. Pix2Pix can convert images between distinct domains A and B . Based on semantic segmentation, we explored the use of the annotated masks in order to train the network to convert real images to masks (and vice versa). After training, we give the generator G a mask, and the generator produces a new image, with the requested new label. We tested several approaches to encode this new label information. Those approaches are illustrated in Figure 6.3 and detailed below (in the order they were conceived):

1. Using different colored semantic masks to encode the dial perspective and check if Pix2Pix generates different pointed values (1 color). Result: it didn't generate the pointers and digits correctly.
2. Using different colored semantic masks to encode the correct label (10 colors) and dial perspective (1 color). Result: Confusion regarding direction, correct position (a smaller pointer was created) and the digits are distorted.
3. Using different colored semantic masks to encode the correct label and dial perspective for only pointed values equal to 9 and only clockwise dials (1 color). Result: Confusion regarding correct position, though more consistent numbers were generated.
4. Using different colored semantic masks to encode the dial perspective and dial direction (2 colors: 1 for CW and the other for CCW). Result: Confusion regarding the pointer generation (no pointer was generated), however more consistent numbers were generated.
5. Using different colored semantic masks to encode the dial perspective and dial direction (2 colors: 1 for CW and the other for CCW) and a different color to encode and approximation of the dial using a line segment (1 color), adding up to only 3 colors. Result: No confusion regarding correct pointer position, and consistent numbers generated (our final and best approach).

We generated 50,000 new synthetic images, using the Approach 5. The process consists of generating individual dials, and afterwards inserting the synthetic dials on real meter images using the annotated masks as a reference to crop them accordingly and the irregular quadrilateral as a reference for perspective. The dial positions were generated using randomly and uniformly distributed fractional labels. We present some samples of the final images in Figure 6.4

We still haven't increased the methods performance in initial experiments. However, there are still evaluations in progress. Our next goal is to generate 200,000 and 500,000 new images in order to evaluate how they could positively impact in the improvement of the methods' performance.

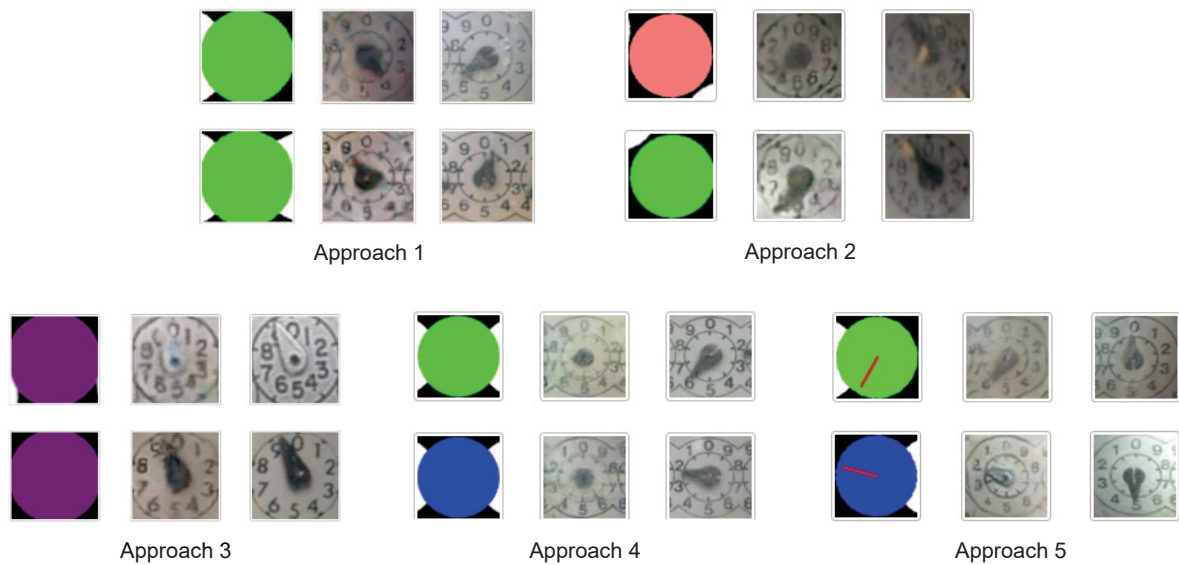


Figure 6.3: For each approach two samples are given (one per line). For each sample: on the left the masks are displayed, on the right the original dial (for reference) and in the middle the synthetic images generated by Pix2Pix.



Figure 6.4: Samples of the 50,000 generated images. After generating the individual dials, using the annotated mask and bounding rectangle, we correct the perspective and insert the synthetic dials in the original image. Afterwards, the image is warped back into the original perspective.

7 CONCLUSIONS

Imaged-based ADMR is a faster and less laborious solution than manual on-site reading, and easier to deploy than the replacement of old meters. In this work, we presented the issues and challenges regarding the automatic reading of dial meters since there are many open challenges in this context.

We came up with an approach for this problem that combines a detection model with a regression model to enable error correction for the most significant dials, thus decreasing the MAE 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.

We introduced a public real-world dataset (shared under request), called UFPR-ADMR-v2, for automatic dial meter reading, that includes 5,000 fully annotated images acquired on site by employees of one of the largest companies of the Brazilian electricity sector (Copel, 2020). The proposed dataset contains a well-defined evaluation protocol, which enables a fair comparison of different methods in future works.

Compared to the UFPR-ADMR-v1 dataset, introduced in our previous work (Salomon et al., 2020), 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 and made possible the use of Pix2Pix to generate a new synthetic dataset containing more than 50,000 dial images.

Considering that the image scenario is challenging in most cases, the deep networks YOLO and Xception achieved promising results. This straightforward approach, without RoI detection or image preprocessing, simplified the traditional AMR pipeline (Vanetti et al., 2012; Gallo et al., 2015; Laroca et al., 2019), reducing the number of steps required to obtain the dial meter readings.

In addition, we combined both networks in order to create a post-processing correction method, that increased performance. We employed several techniques to undertake many sources of errors such as: severe rotation, uneven lighting and neighbor values led us to a robust system that can perform well in unconstrained environments.

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 ($MAE = 1$) at a derisive cost (less than R\$1,00), we could increase the system’s MRR from 92.50% to 98.90%. A cost that would almost certainly be compensated in a subsequent reading.

7.1 FUTURE WORK

There is a lot of room for improvement, such as new methods to address the boundaries issues between the markers, which should solve most of the errors. Also, the development of a YOLO-based detection model that can perform regression would enable a faster (and possibly with a better performance) reading system. In addition, a new loss function that penalizes errors on the leftmost dials should help to reduce the absolute error (minimizing the absolute error is of paramount importance to the service company). The use of GANs could enable a better per-class balancing, removing biases in order to increase the system’s performance, impacting specially the performance of segmentation-free approaches, simplifying the ADMR pipeline.

7.2 PUBLICATIONS

The following articles were accepted and published. The first and the last are the result of the research related to the current subject ADMR. The other two articles have contributed indirectly for obtaining experience and consolidating knowledge on deep learning research, experimental protocols and developing academic writing skills.

- 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.
- G. Salomon, A. S. Britto Jr., R. Vareto, W. R. Schwartz, D. Menotti, “Open-set Face Recognition for Small Galleries Using Siamese Networks” in *International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020.
- D. V. Ruiz, G. Salomon, E. Todt, “Can Giraffes Become Birds? An Evaluation of Image-to-image Translation for Data Generation” in *Computer on the Beach*, 2020.
- G. Salomon, R. Laroca, D. Menotti, “Image-based Automatic Dial Meter Reading in Unconstrained Scenarios” (*To be submitted*) in (*To be defined*), 2021.

REFERENCES

- ANEEL (2019). Perdas de energia elétrica na distribuição. Available in: <https://www.aneel.gov.br/documents/654800/18766993/Relat%C3%B3rio+Perdas+de+Energia+2019.pdf/6cb0bf36-4074-bbc3-d15d-ed370f44b34b>. Accessed in: 16-06-2021.
- Atienza, R. (2021). Vision transformer for fast and efficient scene text recognition. In *International Conference on Document Analysis and Recognition (ICDAR)*.
- Ausgrid (2020). Types of meters. Available in: <https://www.ausgrid.com.au/Your-energy-use/Meters/Type-of-meters>. Accessed in: 12-02-2020.
- Baek, J. et al. (2019). What is wrong with scene text recognition model comparisons? Dataset and model analysis. In *International Conference on Computer Vision (ICCV)*.
- Bao, Tan, Liu, and Miao (2019). Computer vision measurement of pointer meter readings based on inverse perspective mapping. *Applied Sciences*, 9(18):3729.
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint*, arXiv:2004.10934:1–14.
- Borisyuk, F., Gordo, A., and Sivakumar, V. (2018). Rosetta: Large scale system for text detection and recognition in images. In *ACM SIGKDD Intl. Conference on Knowledge Discovery & Data Mining*, page 71–79.
- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11(2).
- Calefati, A., Gallo, I., and Nawaz, S. (2019). Reading meter numbers in the wild. In *Digital Image Computing: Techniques and Applications*, pages 1–6.
- Callmepower (2020). Types of electricity meters. Available in: <https://callmepower.com/useful-information/electricity-meter-types>. Accessed in: 12-02-2020.
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1251–1258.
- Copel (2020). Energy Company Of Paraná. Available in: <http://www.copel.com/hpcopel/english/>. Accessed in: 12-02-2020.
- da Silva Marques, R. C., Serra, A. C., França, J. V. F., Diniz, J. O. B., Braz, G., de Almeida, J. D. S., da Silva, M. I. A., and Monteiro, E. M. G. (2019). Image-based electric consumption recognition via multi-task learning. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 419–424. IEEE.
- Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International journal of computer vision*, 88(2):303–338.
- Fang, Y., Dai, Y., He, G., and Qi, D. (2019). A mask RCNN based automatic reading method for pointer meter. In *Chinese Control Conference*, pages 8466–8471.

- Gallo, I., Zamberletti, A., and Noce, L. (2015). Robust angle invariant GAS meter reading. In *International Conference on Digital Image Computing: Techniques and Applications*, pages 1–7.
- Gao, J., Guo, L., Lv, Y., Wu, Q., and Mu, D. (2018). Research on algorithm of pointer instrument recognition and reading based on the location of the rotation center of the pointer. In *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 1405–1410.
- Gonçalves, G. R., Diniz, M. A., Laroca, R., Menotti, D., and Schwartz, W. R. (2018). Real-time automatic license plate recognition through deep multi-task networks. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 110–117.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Gómez, L., Rusiñol, M., and Karatzas, D. (2018). Cutting sayre’s knot: Reading scene text without segmentation. Application to utility meters. In *IAPR Intern. Workshop on Document Analysis Systems*, pages 97–102.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.
- He, P., Zuo, L., Zhang, C., and Zhang, Z. (2019). A value recognition algorithm for pointer meter based on improved Mask-RCNN. In *International Conference on Information Science and Technology*, pages 108–113.
- Howells, B., Charles, J., and Cipolla, R. (2021). Real-time analogue gauge transcription on mobile phone. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 2369–2377.
- Huang, Y., Dai, X., and Meng, Q. (2019). An automatic detection and recognition method for pointer-type meters in natural gas stations. In *Chinese Control Conference*, pages 7866–7871.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.
- Jiale, H., En, L., Bingjie, T., and Ming, L. (2011). Reading recognition method of analog measuring instruments based on improved hough transform. In *International Conference on Electronic Measurement Instruments*, volume 3, pages 337–340.
- Kabalci, Y. (2016). A survey on smart metering and smart grid communication. *Renewable and Sustainable Energy Reviews*, 57:302–318.
- Kanagarathinam, K. and Sekar, K. (2019). Text detection and recognition in raw image dataset of seven segment digital energy meter display. *Energy Reports*, 5:842–852.
- Laroca, R., Araujo, A. B., Zanlorensi, L. A., de Almeida, E. C., and Menotti, D. (2021a). Towards image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach. *IEEE Access*, 9:67569–67584.
- Laroca, R., Barroso, V., Diniz, M. A., Gonçalves, G. R., Schwartz, W. R., and Menotti, D. (2019). Convolutional neural networks for automatic meter reading. *Journal of Electronic Imaging*, 28:1–14.

- Laroca, R., Zanlorensi, L. A., Gonçalves, G. R., Todt, E., Schwartz, W. R., and Menotti, D. (2021b). An efficient and layout-independent automatic license plate recognition system based on the YOLO detector. *IET Intelligent Transport Systems*, 15(4):483–503.
- Lauridsen, J. S., Grassmé, J. A. G., Pedersen, M., Jensen, D. G., Andersen, S. H., and Moeslund, T. B. (2019). Reading circular analogue gauges using digital image processing. In *VISIGRAPP*.
- Lee, C. and Osindero, S. (2016). Recursive recurrent nets with attention modeling for OCR in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2231–2239.
- Li, C., Su, Y., Yuan, R., Chu, D., and Zhu, J. (2019). Light-weight spliced convolution network-based automatic water meter reading in smart city. *IEEE Access*, 7:174359–174367.
- Liao, S., Zhou, P., Wang, L., and Su, S. (2019). Reading digital numbers of water meter with deep learning based object detector. In *Chinese Conference on Pattern Recognition and Computer Vision*, pages 38–49.
- Liu, L. et al. (2019a). Deep learning for generic object detection: A survey. *International Journal of Computer Vision*.
- Liu, W., Chen, C., Kwan-Yee K. Wong, Z. S., and Han, J. (2016a). STAR-Net: A spatial attention residue network for scene text recognition. In *British Machine Vision Conference (BMVC)*, pages 1–13.
- Liu, W. et al. (2016b). SSD: Single shot multibox detector. In *European Conference on Computer Vision (ECCV)*, pages 21–37.
- Liu, Y., Liu, J., and Ke, Y. (2019b). A detection and recognition system of pointer meters in substations based on computer vision. *Measurement*, 152:107333.
- Mai, X., Li, W., Huang, Y., and Yang, Y. (2018). An automatic meter reading method based on one-dimensional measuring curve mapping. In *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*, pages 69–73.
- Navarro, G. (2001). A guided tour to approximate string matching. *ACM computing surveys (CSUR)*, 33(1):31–88.
- Nodari, A. and Gallo, I. (2011). A multi-neural network approach to image detection and segmentation of gas meter counter. In *IAPR Conference on Machine Vision Applications (MVA)*, pages 239–242.
- Ocampo-Vega, R. et al. (2013). Image processing for automatic reading of electro-mechanical utility meters. In *Mexican International Conference on Artificial Intelligence*, pages 164–170.
- Powers, D. M. (2015). What the f-measure doesn't measure: Features, flaws, fallacies and fixes. *arXiv preprint arXiv:1503.06410*.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788.
- Redmon, J. and Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525.

- Redmon, J. and Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint*, arXiv:1804.02767.
- Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.
- Ren, Y., Wang, Z., Xu, W., and Chen, J. (2019). A meter pointer detection method based on window function scanning. In *Proceedings of the 2019 4th International Conference on Robotics, Control and Automation*, pages 38–42.
- Salomon, G., Laroca, R., and Menotti, D. (2020). Deep learning for image-based automatic dial meter reading: Dataset and baselines. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- Shi, B., Bai, X., and Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11):2298–2304.
- Shi, B., Bai, X., and Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11):2298–2304.
- Shi, B., Wang, X., Lyu, P., Yao, C., and Bai, X. (2016). Robust scene text recognition with automatic rectification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4168–4176.
- Silva, S. M. and Jung, C. R. (2017a). Real-time brazilian license plate detection and recognition using deep convolutional neural networks. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 55–62.
- Silva, S. M. and Jung, C. R. (2017b). Real-time brazilian license plate detection and recognition using deep convolutional neural networks. In *Conference on Graphics, Patterns and Images*, pages 55–62.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*, pages 1–12.
- Smith, R. (2007). An overview of the tesseract ocr engine. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pages 629–633.
- Souare, M., Merat, F., and Papachristou, C. (2014). Efficient way of reading rotary dial utility meter using image processing. In *ASEE*.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.
- Tang, Y., Ten, C., Wang, C., and Parker, G. (2015). Extraction of energy information from analog meters using image processing. *IEEE Transactions on Smart Grid*, 6(4):2032–2040.
- Tsai, C., Shou, T. D., Chen, S., and Hsieh, J. (2019). Use SSD to detect the digital region in electricity meter. In *International Conference on Machine Learning and Cybernetics (ICMLC)*, pages 1–7.

- U.S. Energy Information Administration (2019). Electric power annual 2018. Available in: <https://www.eia.gov/electricity/annual/pdf/epa.pdf>. Accessed in: 12-02-2020.
- Vanetti, M., Gallo, I., and Nodari, A. (2012). Gas meter reading from real world images using a multi-net system. *Pattern Recognition Letters*, 34(5):519–526.
- Wang, J. and Hu, X. (2017). Gated recurrent convolution neural network for OCR. In *Conference on Neural Information Processing Systems*.
- Yang, F., Jin, L., Lai, S., Gao, X., and Li, Z. (2019). Fully convolutional sequence recognition network for water meter number reading. *IEEE Access*, 7:11679–11687.
- Zhang, L., Fang, B., Zhao, X., and Zhang, H. (2016). Pointer-type meter automatic reading from complex environment based on visual saliency. In *International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, pages 264–269.
- Zhao, S., Li, B., Yuan, J., and Cui, G. (2005a). Research on remote meter automatic reading based on computer vision. In *2005 IEEE/PES Transmission Distribution Conference Exposition: Asia and Pacific*, pages 1–4.
- Zhao, S., Li, B., Yuan, J., and Zhao, D. (2005b). Key elements extraction based on particle feature and rbfn in new meter calibration method. In *2005 IEEE International Conference on Industrial Technology*, pages 795–798. IEEE.
- Zheng, W., Yin, H., Wang, A., Fu, P., and Liu, B. (2017). Development of an automatic reading method and software for pointer instruments. In *International Conference on Electronics Instrumentation Information Systems*, pages 1–6.
- Zhu, J., Huang, W., Chen, W., and Zhang, X. (2019). An ellipse fitting with pso for automatic reading recognition of pointer instruments. In *2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS)*, pages 42–46.
- Zuo, L., He, P., Zhang, C., and Zhang, Z. (2020). A robust approach to reading recognition of pointer meters based on improved mask-rcnn. *Neurocomputing*.