IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 January 2025, accepted 20 February 2025, date of publication 3 March 2025, date of current version 12 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3547225

RESEARCH ARTICLE

Water Meter Reading Based on Text Recognition Techniques and Deep Learning

BAY NGUYEN VAN[1], ANH NGUYEN[2], KIET TRAN-TRUNG[1], THIEN HO HUONG[1], HA DUONG THI HONG[1], HAU NGUYEN TRUNG[1], AND VINH TRUONG HOANG[1]

[1]Faculty of Information Technology, Ho Chi Minh City Open University, Ho Chi Minh City 70000, Vietnam
[2]FPT School of Business and Technology (FSB), FPT University, Ho Chi Minh City 71216, Vietnam

Corresponding author: Vinh Truong Hoang (vinh.th@ou.edu.vn)

This work was supported by Ho Chi Minh City Open University under Grant 01/E2023.05.2/HĐ-HTQLKH and Grant E2023.05.2.

ABSTRACT In an increasingly technology-driven and automated world, optimizing energy management efficiency is crucial. A key component of this optimization is the precise measurement of water consumption, which plays a vital role in the allocation of resources for both households and businesses. While automated water meters are available, their adoption remains limited, particularly in less developed regions where manual meter reading is still common. This manual process often leads to inaccuracies and inefficiencies. This research aims to automate water meter readings, particularly in challenging and diverse environmental conditions. We propose a structured two-step approach to Automated Meter Reading, involving the identification of the counter region followed by digit segmentation and recognition. Using Convolutional Neural Networks, particularly YOLOv8 for detecting the counter region, we evaluate six CNN-based techniques for digit segmentation and recognition: PP-OCRv3, Structure-Preserving Inner Offset Network, Transformer-based Optical Character Recognition, RobustScanner, Show-Attend-and-Read, and Convolutional Recurrent Neural Network.

INDEX TERMS Automated meter reading, energy management, consumption data collection, digit recognition, deep learning, OCR, optical character recognition.

I. INTRODUCTION

In an increasingly technology-driven and automated world, improving energy management efficiency has become essential for saving time and optimizing resources [1]. One critical aspect of this is accurately measuring water consumption, which is vital for determining the amount of water used by households and businesses [2]. Automated water meter reading refers to the process of automatically recording water consumption levels. Although smart electronic meters are available, they are not widely implemented in many countries, particularly in less developed regions [3], [4]. As a result, manual meter reading by personnel remains common. In this method, employees physically collect data from each meter and manually enter the readings into the system. This manual approach is prone to errors and often necessitates on-site verification, leading to costly, time-consuming, and labor-intensive offline checks [5].

Developing a reliable automatic reading system for water meters is challenging due to the diverse imaging conditions and the often-difficult environments in which these meters are located. As shown in Figure 1, real-world meter images exhibit a variety of challenges that must be addressed.

First, images exhibit significant variations in lighting, resolution, and background. Second, the positions and rotation angles of the meters are often unpredictable. Finally, low image resolution and dust accumulation on meters further complicate accurate reading of meter values.

Implementing automated water meter reading can significantly reduce human-induced errors and save workforce resources [5], [6]. Moreover, the reading process can be fully automated by utilizing cameras within the meters. Automatic meter reading offers advantages such as lower costs and quick

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenbao Liu.

FIGURE 1. Water meters in real-life challenges.

installation, as it does not necessitate the replacement or upgrading of existing meters [7].

This research is vital despite existing studies in the Automated Meter Reading (AMR) field for several reasons. Firstly, the limited adoption of smart meters, especially in less developed regions, makes developing methods for automatic reading from traditional meters practical. Secondly, the integration of new technologies like artificial intelligence and image processing provides opportunities to handle diverse environmental conditions, improving accuracy and efficiency. Thirdly, having accurate water consumption data benefits both households and energy companies, enabling efficient resource planning and reducing waste. Lastly, the developed methods can be scaled for various meter types and usage contexts, contributing to global resource conservation. In conclusion, this research addresses challenges, leverages technology, and enhances energy management, particularly relevant for places like Vietnam. This study leverages several contemporary deep learning technologies for object detection and text recognition:

- **You Only Look Once (YOLO)**: YOLO is a real-time object detection model that processes the entire image at once and predicts bounding boxes efficiently. Its latest version, YOLOv8, improves accuracy and computational efficiency, making it suitable for counter region detection in water meter images.
- **Convolutional Neural Networks (CNNs)**: CNNs are widely used for image processing tasks due to their ability to learn spatial hierarchies of features. In this study, CNNs are utilized for detecting and segmenting numerical regions and recognizing digits.
- **Recurrent Neural Networks (RNN)**: RNNs are commonly employed for sequential data processing. They are integrated with OCR models to recognize and process numeric sequences extracted from the meter displays. These technologies form the backbone of our proposed system, combining detection (YOLOv8) and recognition (OCR models) to achieve automated water meter reading.

Furthermore, we introduce a structured two-step strategy specifically designed for AMR. The first stage focuses on identifying the counter region, followed by a second phase centered on digit segmentation and recognition. To achieve this, we leverage the powerful capabilities of CNNs [8]. For detecting the counter region, we utilize YOLOv8 [9], a specialized variant of the YOLO object detector, known for its efficiency and accuracy in object detection tasks.

Following the counter region detection, we perform an in-depth evaluation of six different CNN-based methodologies: PP-OCRv3 [10], Structure-Preserving Inner Offset Network [11], **Tr**ansformer-based **O**ptical **C**haracter **R**ecognition (TrOCR) [12], RobustScanner [13], **S**how-**A**ttend-and-**R**ead (SAR) [14], and **C**onvolutional **R**ecurrent **N**eural **N**etwork (CRNN) [15]. These methodologies are critically assessed for their performance in digit segmentation and recognition within the context of AMR. Our selection of these approaches is based on their demonstrated effectiveness in similar applications, making them strong candidates for this evaluation within the AMR framework.

This paper brings valuable contributions:

- We created a vast dataset for water meters, consisting of 2172 training images and 1500 test images. These images cover various challenging environments, enhancing the reliability of automatic readings.
- All images underwent manual annotations to establish their accurate ground truths. Annotations were made for both the meter displays and individual digits.
- We introduced a robust model using CNNs to automatically read structured water meter instruments.
- Through extensive experiments, we demonstrated that our method effectively handles diverse environmental conditions, ensuring a dependable meter reading performance.

The rest of the research paper is organized as follows. Section II looks at existing studies in the field, Section III details our experiment results and analyses of various models, and Section IV concludes with our study's key findings and suggestions for future research.

II. RELATED WORKS

The advancement of smart grid technology has significantly transformed the energy sector by enabling more efficient energy management and real-time consumption monitoring. A key element of this transformation is the accurate and automated reading of meters. Traditional manual meter reading is not only labor-intensive and prone to errors but also lacks the capability for real-time data collection. To overcome these limitations, researchers have increasingly turned to deep learning techniques to develop automated systems that can accurately read meters from images captured in real-world environments. These systems offer a promising solution to the challenges associated with manual meter reading and play a vital role in the modernization of energy infrastructures.

Salomon et al. [16] presented a method in 2020 for automatically reading dial displays on electricity, water, and gas meters using images. They created a dataset called UFPR-ADMR with 2,000 images taken in real-world conditions by employees of the Brazilian electricity company Copel. The authors used deep learning techniques, specifically Faster R-CNN, an improvement upon the Region-based Convolutional Neural Network (R-CNN) and YOLOv3, for detecting and recognizing the digits on the dial displays. The results showed that both deep learning models performed well, achieving a 100.0% F1-score for dial detection and 93.6% for recognizing the individual digits on the dials. However, there were some challenges. The recognition rate for entire meters was 75.25%, indicating some misidentification issues. Additionally, the models struggled with challenging image conditions, like low light, dirt, and unusual angles, leading to a decrease in dial recognition accuracy. In 2022 [17] Salomon et al. presented a method for automatically reading electricity meter dials from images in unconstrained situations. The authors proposed a combination of detection and regression models to reduce reading errors. They trained and tested 18 deep models using various frameworks such as Darknet, Keras, and PyTorch, all on the same dataset. The dataset used in the study was called UFPR-ADMR-v2, containing over 5000-meter images collected in real-world scenarios, with fully annotated meter readings. Experimental results showed that object detection-based models (e.g., YOLOv3 and YOLOv4) outperformed regression and segmentation-free models. The best results were achieved by combining the top-performing object detection model (YOLOv4) with the AngReg regression model based on the Xception architecture. However, the paper also highlighted opportunities for further improvements, particularly in handling images with poor quality and unfavorable lighting conditions.

Martinelli et al. introduced an automated method for reading water meters using the YOLOv5s deep learning model in a smart grid monitoring system [18]. The objective was to identify and classify the digits on water meters from images. To evaluate the approach, the researchers utilized a dataset comprising 1000 water meter images and augmented the data using image augmentation techniques. The experimental results demonstrated that the method performed well in automatically reading water meters. However, it also had some limitations, such as possible confusion between certain digits and reliance on image quality. To address these issues, further research and development are needed to ensure accuracy across various types of water meters and under different conditions.

Li et al. introduced an innovative method for automating water meter reading through image recognition [19]. Their approach leveraged a lightweight spliced convolutional network, specifically designed to streamline computation and minimize the model's size, making it more efficient for practical deployment. The method was rigorously tested on a real-world dataset comprising 6,000 images, where it exhibited high accuracy in recognizing water meter numbers. While the method showed promising results, it faced challenges when processing meters with long numerical sequences, which complicated digit segmentation. To address these issues, the authors proposed further research aimed at optimizing the network's architecture and improving its performance in handling extended numerical sequences. Future work would also focus on enhancing the method's robustness and scalability, enabling it to manage a broader range of water meter formats and environmental conditions. This expansion could potentially involve refining the segmentation algorithm, incorporating more advanced deep learning techniques, or exploring hybrid models to overcome the difficulties posed by longer meter numbers and ensure reliable performance across diverse scenarios.

Imran et al. [20] implemented the YOLOv3 deep learning algorithm to automate the reading of energy meters from images. Their study utilized a comprehensive dataset of 10,000 electrical energy meter images, applying data augmentation techniques to enhance the model's ability to recognize digits under various conditions, such as occlusion, changes in scale, and different orientations. The results indicated that YOLOv3 significantly outperformed traditional approaches, achieving an impressive 98% average recognition accuracy across all digits. The proposed method demonstrated considerable potential in reducing the time and effort required for manual meter readings, while also facilitating the automatic recognition of serial numbers, an essential feature for automating billing processes. Despite its success, the paper identified the need for further improvements, such as expanding the dataset and optimizing the model to achieve even greater accuracy in real-world scenarios. Future research may focus on integrating YOLOv3 with IoT devices and developing a graphical user interface for real-time deployment, paving the way for a fully automated and efficient meter reading system.

Hong et al. developed an automated system for reading water meter digits from images, specifically designed to perform well in challenging environments such as low light, blurry images, and dirty conditions [21]. The method consisted of four key steps: first, the water meter's location within the image was detected using a YOLOv3-based model. Next, a deep learning approach was employed to adjust the image orientation to ensure accurate readings. Following this, another YOLOv3-based model was used to precisely locate the digits on the water meter. Finally, a deep learning model was applied to read and interpret the digit values.

The system was trained using a large dataset of 9,500 training images and 500 test images collected from diverse environments, which helped to enhance the system's adaptability and robustness. Experimental results demonstrated that the proposed method achieved high accuracy in reading water meter digits, even under difficult conditions. However, the system encountered limitations when dealing with images obstructed by foreign objects, leading to occasional errors in digit recognition. Further optimization is needed to improve the system's reliability and accuracy when faced with such obstructions.

Laroca et al. proposed an AMR approach using CNNs for the detection and recognition of electric meter counters in images [22]. The study primarily utilized the UFPR-AMR dataset, a newly released public dataset specifically designed for this purpose. The paper detailed the methodology, models, and results, highlighting the use of Fast-YOLO for counter detection and three CNN-based models—CR-NET, Multi-Task Learning, and CRNN—for counter recognition. To enhance performance, data augmentation techniques were applied during the training process. The UFPR-AMR dataset, consisting of 2,000 images with manually labeled digits, played a central role in the experimentation. Among the models evaluated, CR-NET performed particularly well, achieving a counter recognition accuracy of 94.13% with data augmentation. The study also extended its evaluation to other public datasets, demonstrating the robustness of the Fast-YOLO model for counter detection across various scenarios. While the proposed approach delivered impressive recognition rates, particularly with CR-NET, the paper also identified several limitations. Common challenges included errors caused by rotating digits and artifacts in the counter regions, such as reflections and dirt, which occasionally led to recognition inaccuracies. These issues underscore the need for further refinement to improve robustness in such adverse conditions

Additionally, dataset bias presents challenges for segmentation-free approaches, negatively affecting recognition rates, particularly for less frequent digits [22]. In a related study, Liang et al. [23] explored the optimization of water meter reading efficiency using deep learning-based image recognition techniques. The study compared the effectiveness of three established models—Faster R-CNN, **S**ingle **S**hot MultiBox **D**etector (SSD), and YOLOv3—in identifying digits within black rectangular boxes on water meter images. Two datasets were employed: one consisting of original water meter images and another of images cropped to focus solely on the black rectangle containing the digits.

The proposed approach involved training the models to recognize either the entire image or just the black rectangular box. In this configuration, YOLOv3 demonstrated superior performance, achieving a recognition accuracy of 90.61%. The paper underscores the potential of image recognition technology to enhance the efficiency of meter reading processes while reducing costs associated with manual readings.

Despite these promising results, the study acknowledges certain limitations, particularly the impact of external factors such as varying illumination conditions and the presence of other numbers on the dial, which can interfere with the digit recognition process. Addressing these challenges is crucial for further improving the accuracy and reliability of AMR systems.

Koščević et al. proposed an innovative approach utilizing Faster R-CNN deep learning for reading various types of residential energy meters [24]. Their method involved modifying the Faster R-CNN model to accurately detect both the counter and serial number regions on meters, employing irregular quadrangle annotations for better precision. To support this approach, the authors developed a manually annotated dataset and implemented a data augmentation system to expand the dataset and streamline the annotation process. The results demonstrated the method's robustness across different meter types, with high success rates in digit detection. The approach proved to be both practical and cost-effective. However, the paper identified some limitations, particularly when dealing with the variability of real-world conditions and unique meter characteristics that were not sufficiently represented in the training dataset. Addressing these challenges is essential for further enhancing the method's adaptability and reliability in diverse environments.

These studies demonstrate significant progress in automating meter reading, highlighting their potential to improve accuracy and efficiency in smart grid systems. However, they also highlight some limitations, particularly in handling images with poor quality and unfavorable lighting conditions. Future research endeavors should focus on addressing these challenges to ensure reliable performance across various scenarios. Additionally, as the deployment of smart meters continues to expand, the integration of these advanced meter reading methods with the Internet of things (IoT) and real-time data processing systems is a promising avenue for further exploration. By overcoming the current limitations, such systems can revolutionize energy distribution and billing processes, ultimately benefiting both service providers and consumers.

III. EXPERIMENTS

A. IMAGE-BASED AUTOMATIC WATER METER READING ARCHITECTURE

Figure 2 illustrates an automated process for reading a water meter's numerical display using computer vision technology. The process starts with an image of the water meter, which is fed into the system. The first computational step involves Object Detection [25], where the algorithm identifies and isolates the section of the image containing the meter's numeric display. This step is analogous to teaching a computer to recognize objects by exposing it to numerous images, like how humans learn to recognize a cat by seeing it in various places and from different angles [26].

Once the numeric display is isolated, the system extracts the numbers into a separate image for processing using **O**ptical **C**haracter **R**ecognition (OCR) technology. OCR converts the text from the image into editable, searchable text data, much like transforming scanned documents into digital text [27]. The final output of this process is the water meter reading in a digital format, such as "00241801," ready for further digital processing or record-keeping

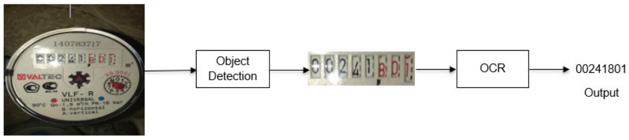


FIGURE 2. Water meter reading process flow.

B. TYPES OF WATER METERS

Figure 3 depicts a typical mechanical water meter. While various manufacturers may produce these meters with slight differences, their displays and reading conventions are generally consistent. This uniformity enables the application of automated methods for reading these meters, based on the segmentation and recognition of the digits. In the image, the value within the gold box represents the "digit reading," illustrating the process of interpreting this type of water meter.

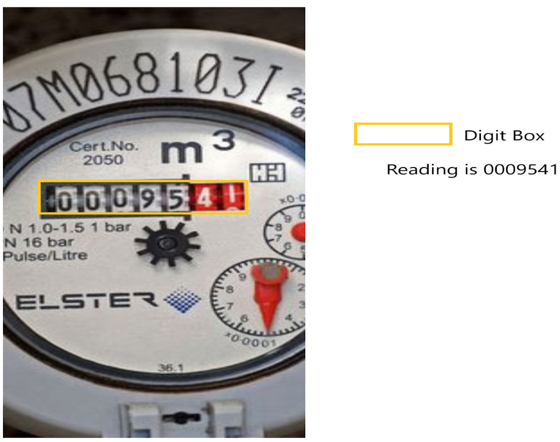


FIGURE 3. Water meter schematic.

Currently, common water meters include 4-digit, 5-digit, 6-digit, 7-digit, and 8-digit meters, as shown in Figure 4. These meters display sequences of numbers in both black and red. The black numbers indicate the volume of water in cubic meters (m^3), while the red numbers represent liters, broken down into hundreds, tens, single units, and tenths of a liter. For these types of meters, recording the water usage is straightforward: simply note the sequence of black numbers from left to right, as they provide the total volume in cubic meters. For example:

- 0004 (4 black digits) can be recorded as 4 m^3.
- 00316 (5 black digits) can be recorded as 316 m^3.
- 03804 (4 black digits, 1 red digit) can be recorded as 380 m^3.
- 0005631(5 black digits, 2 red digits) can be recorded as 56 m^3.
- 0007337 (4 black digits, 3 red digits) can be recorded as 7 m^3.
- 00455552 (4 black digits, 4 red digits) can be recorded as 45 m^3.

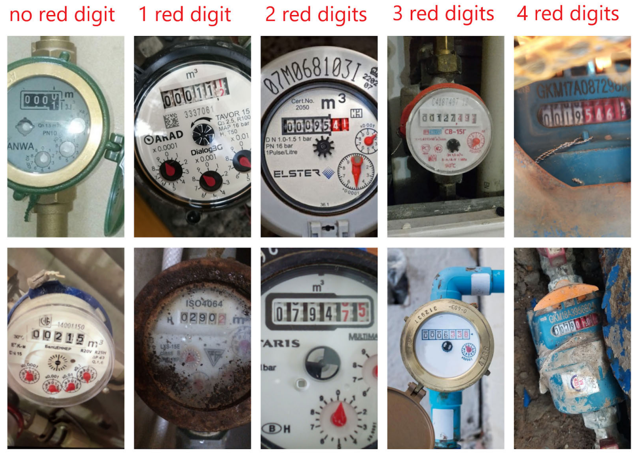


FIGURE 4. Types of water meters.

C. DATASET

1) ORIGIN AND DATA COLLECTION

The dataset comprises 3,672 images collected from diverse sources:

Yandextoloka dataset (1,244 images): Images of mechanical water meters with 8 digits obtained from the Yandextoloka Water Meters dataset on Kaggle (see Figure 5).

FIGURE 5. Yandextoloka dataset.

Mì AI dataset (1,600 images): Images of mechanical electricity meters with 5 digits sourced from the Mì AI website (see Figure 6).

Directly Gathered Data from the Community (828 images): Mainly images of mechanical water and mechanical electricity meters with 5 to 6 digits, collected personally through permission and billing from the local community (see Figure 7).

The photos vary in quality and lighting, some taken during the day and others at night. We have chosen this dataset, consisting of 3672 images collected from various sources, because it directly aligns with our thesis's goal of automating the reading of water meter numbers. The

FIGURE 6. Mì AI dataset.

FIGURE 7. Self-collected dataset.

dataset comprises images of both water and electricity meters, offering a valuable opportunity to develop a comprehensive meter-reading solution. The data's reliability from reputable sources, community engagement, scalability, and direct relevance to real-world applications make it the ideal choice for our thesis focused on reading water meter numbers. This dataset equips us with the necessary tools to create an efficient and accurate automated meter-reading solution for water meters.

2) DATA PREPROCESSING

Filter out low-quality or irrelevant images: Identify and eliminate images that do not meet quality standards or do not contain relevant information about mechanical electricity or mechanical water meters.

Uniform Image Size: All images have been resized to 640 × 480 pixels to ensure uniformity in processing and reduce computational complexity.

Isolate the portion containing the measurement digits on electricity or water meters: Segment and extract the area of the image that specifically contains the numerical measurements from the original image.

3) SELF-LABELING AND DATA SPLIT

Self-Labeling: The labeling process involved personal annotation to identify meter readings on water and electricity meters. This approach ensures hands-on accuracy and control over the labeled data.

Training and Testing Data Split: The dataset has been split into two sets:

- Training Set: Consisting of 2172 images.
- Testing Set: Consisting of 1500 images.

Splitting the dataset aids in evaluating the model's performance on unseen data, retaining a sufficiently large portion to ensure the representativeness of the testing set.

Figure 8 demonstrates the labeling process involved annotating all numeric values that indicate water and electricity consumption on the meters.

FIGURE 8. Annotation of the amount of electricity and water consumption.

4) KEY FEATURES OF DATA

Each image contains detailed information about the meter readings, and the preprocessing steps help prepare the data for the model training process. The self-labeling approach provides a personalized touch to the annotation process, ensuring precision and control.

D. EVALUATION METRICS

Evaluating the effectiveness of a model is a critical component of the training process, as it helps determine how well the model performs on given tasks. Various evaluation criteria are employed to assess the model's effectiveness, including several performance metrics such as Accuracy, F1-score, ROC, AUC, Precision, and Recall. For the purpose of this thesis, the Accuracy metric will be used to gauge the model's effectiveness.

Accuracy reflects the ratio of correct predictions made by the model on the test data. It is calculated based on four key parameters: TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). The formula for calculating accuracy is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.1}$$

where:

- TP: Instances correctly classified as positive by the model.
- TN: Instances correctly classified as negative by the model.
- FP: Instances incorrectly classified as positive by the model.
- FN: Instances incorrectly classified as negative by the model.

E. EXPERIMENTAL RESULTS

1) OBJECT DETECTION

The YOLO architecture, inspired by the design of GoogleNet [28], consists of 24 convolutional layers that process and extract features from the input data, 4 layers dedicated to max pooling to reduce dimensionality, and 2 fully connected layers that consolidate the information for final predictions. This structure, as illustrated in Figure 9, enables YOLO to perform efficient object detection by combining feature extraction, spatial reduction, and decision-making in a streamlined architecture.

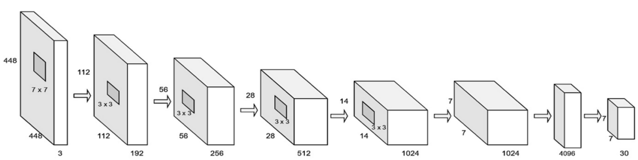


FIGURE 9. Overview of the YOLOv8 architecture, showing convolutional layers, max pooling operations, and fully connected layers for object detection [9].

In January 2023, Ultralytics released YOLOv8 [29], the latest in their YOLO series, with models ranging from tiny to extra-large (YOLOv8n to YOLOv8x). This versatile version excels in object detection, image segmentation, pose estimation, motion tracking, and image classification. Tested with the COCO dataset [30], YOLOv8 outperforms its predecessors across all sizes. The integration of the YOLOv8 object detection model and OCR frameworks was achieved in a sequential pipeline. Specifically:

1. **YOLOv8 Model**:
   - Input: Full water meter images resized to 640 × 640 pixels.
   - Output: Bounding boxes around the numeric display regions, returned as coordinates (x, y, width, height).
   - Postprocessing: The detected bounding boxes were cropped and resized to 640 × 480 pixels for uniformity across OCR models.
2. **OCR Models**:
   - Each cropped bounding box served as input to the respective OCR model.
   - Models such as PP-OCRv3, TrOCR, and RobustScanner were fine-tuned on the water meter dataset to maximize accuracy.
   - Output: Numeric predictions in string format (e.g., '0005631').

A simple post-check validated the predicted numbers for consistency (e.g., removing invalid characters and ensuring numeric-only output). We apply YOLOv8 to detect digit boxes as described in Figure 10.

FIGURE 10. Digit boxes detection.

Figures 11, 12 and 13 represent the results obtained after training with YOLOv8s, YOLOv8m, and YOLOv8x, respectively, over 100 epochs. All images were resized to a uniform resolution of 640 × 480 pixels.

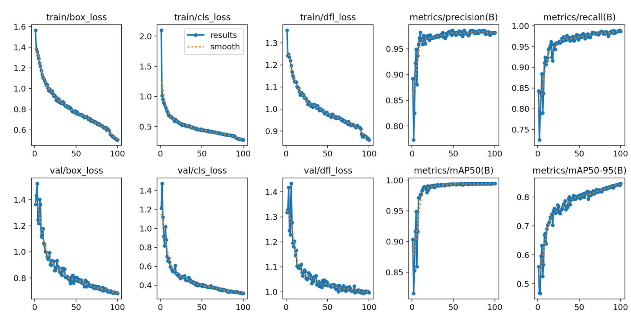


FIGURE 11. Performance results of YOLOv8s after training.

Based on the model training results and with the goal of choosing the most accurate model:

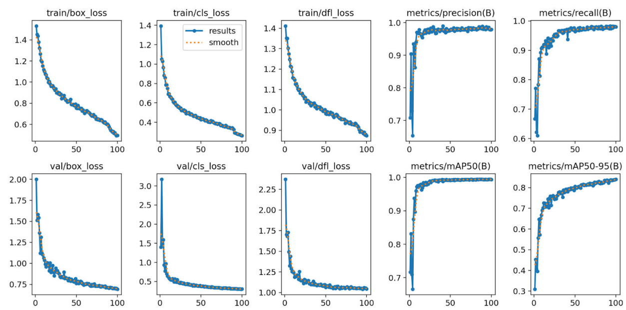


FIGURE 12. **Performance results of YOLOv8m after training.**

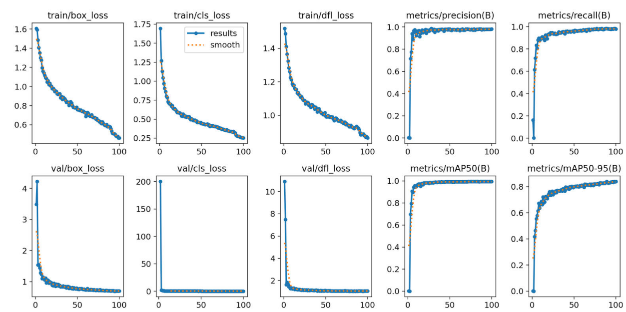


FIGURE 13. **Performance results of YOLOv8x after training.**

Box Loss:

YOLOv8s: Shows a steady decrease in box loss for both training and validation datasets, indicating good learning and generalization.

YOLOv8m: Like YOLOv8s, the box loss decreases steadily, but it tends to be slightly higher on the validation set.

YOLOv8x: Exhibits significantly higher box loss on the validation set, which may suggest overfitting or poorer generalization.

Classification Loss and **Distribution Focal Loss (DFL):**

YOLOv8s: Both metrics decrease steadily without significant fluctuations, indicating accurate object classification and localization.

YOLOv8m: Like YOLOv8s but with slight fluctuations in the validation set, which may indicate some instability in classification and localization.

YOLOv8x: Shows more significant fluctuations and higher loss on the validation set, another sign of potential overfitting.

Precision and Recall:

YOLOv8s: Achieves high precision and recall, indicating the model accurately predicts and doesn't miss objects.

YOLOv8m: Results are similar to YOLOv8s but may not be as stable.

YOLOv8x: Lower than the other two models, indicating poorer prediction capabilities.

mAP:

YOLOv8s: A steady and high mAP increase indicates good object detection capability.

YOLOv8m: Though there are slight fluctuations, it generally also shows a steady and high mAP.

YOLOv8x: Starts with a lower mAP and increases more slowly, a sign of poorer detection performance.

Conclusion and Recommendation:

YOLOv8m demonstrated the highest accuracy based on mAP and other metrics. Although it showed slight fluctuations compared to YOLOv8s, its ability to handle complex situations and extract richer features suggests it can deliver superior results on challenging datasets.

2) OPTICAL CHARACTER RECOGNITION

In this section, our selected models consist of PP-OCRv3 and SPIN from PaddleOCR, along with RobustScanner, SAR, and CRNN from MMOCR, and TrOCR.

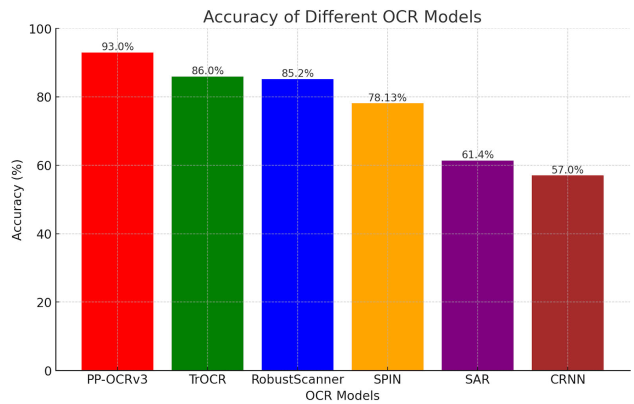


FIGURE 14. **Accuracy of different OCR models.**

As shown in Figure 14, the PP-OCRv3 model from PaddleOCR achieves an impressive 93% accuracy through its three-part structure: character detection, character recognition, and image optimization modules. This architecture likely contributes to its superior performance, efficiently handling image variations and character anomalies – a critical aspect for accurately reading diverse water meter displays. Contrastingly, TrOCR, with 86% accuracy, merges the robust capabilities of Transformers for image processing with RNNs for sequential character recognition. This unique combination is adept at handling continuous character strings, crucial for interpreting the sequential digits in water meter readings. **RobustScanner** demonstrated strong performance (85.2%) in challenging conditions such as low-light images, thanks to its attention mechanism and dynamic feature fusion. Models like **SAR** and **CRNN** underperformed due to their simpler architectures, which lack advanced attention mechanisms to focus on distorted or noisy text. However, its slightly lower accuracy compared to PP-OCRv3 might stem from the challenges in fine-tuning Transformer models for the specific nuances of water meter digits. RobustScanner from MMOCR, at 85.2% accuracy, is tailored to perform under challenging conditions. It incorporates attention mechanisms and a blend of convolutional and recurrent networks.

This design allows it to effectively tackle issues like poor lighting or distorted characters, common in outdoor or hard-to-access water meter environments. The SPIN model from PaddleOCR, though lower in accuracy at 78.13%, might utilize a variation of the architecture seen in PP-OCRv3 but perhaps with less emphasis on advanced image preprocessing, which could account for its reduced effectiveness in deciphering complex meter readings. Lastly, SAR and CRNN from MMOCR, with accuracy rates of 61.4% and 57% respectively, represent more basic OCR models. Employing combinations of CNN and RNN, they are potentially less equipped to handle the intricacies and environmental challenges of water meter reading, as evidenced by their lower accuracy scores. Their simpler architectures might struggle with the fine details and anomalies often encountered in such specialized tasks. In summary, the comprehensive three-part structure, high accuracy, and efficient handling of image variations and anomalies make PP-OCRv3 a strong candidate for accurate and reliable water meter reading applications.

The comparison of the architectures:

PP-OCRv3: Features a lightweight design with a backbone for feature extraction, a neck for feature fusion, and a head for detection, all integrated with a fast and accurate recognition module. Figure 15 shows its architecture. PP-OCRv3 incorporates several key methods for advanced text recognition: *SVTR_LCNet*represents a lightweight text recognition network formed by merging elements from SVTR, a Transformer-based network [31], and the streamlined CNN-based network PP-LCNet [32], *Guided Training of CTC (GTC)*combines attention with Connectionist Temporal Classification (CTC) for accuracy and speed [33], *TextConAug*augments data for contextual richness, the core concept is derived from the ConCLR paper [34], *TextRotNet*pre-trains for better model convergence, this model builds upon the earlier research presented in STR-Fewer-Labels [35], *Unified-Deep Mutual Learning (U-DML)* merges multiple modules for enhanced learning, and *Unlabeled Images Mining (UIM)*generates pseudo-labels from unlabeled images to improve training effectiveness.

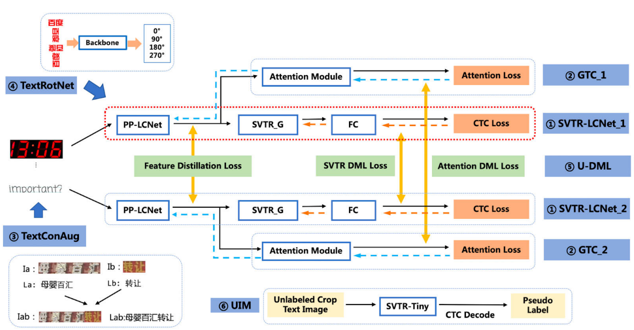


FIGURE 15. PP-OCRv3 recognition model: Framework and Training [10].

SPIN: Employs a structure-preserving network to maintain the geometric structure of the text and an auxiliary network to refine the feature map, potentially reducing its effectiveness if the text's geometric distortions don't align with the training data. It's architecture as illustrated in Figure 16.

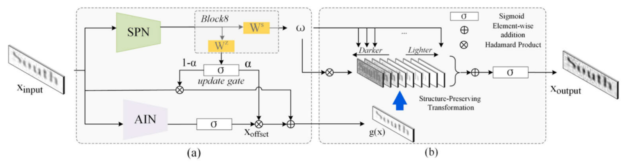


FIGURE 16. The fundamental structure of SPIN [11].

TrOCR: Utilizes the Transformer architecture, known for its self-attention mechanism, allowing it to capture global dependencies within the data. It consists of an image Transformer as an encoder to process visual features and a sequence Transformer as a decoder for character prediction as described in Figure 17.

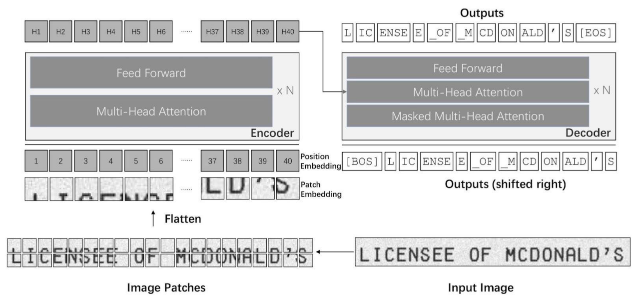


FIGURE 17. The architecture of TrOCR [12].

RobustScanner: Incorporates a visual model to extract features, a sequence model to decode character sequences, and a unique dynamic fusion module to balance the use of visual features and sequence context, which may contribute to its moderate accuracy as shown in Figure 18.

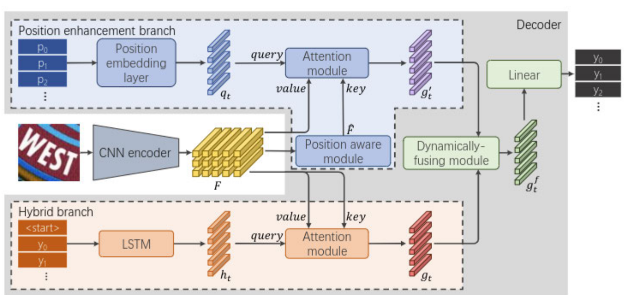


FIGURE 18. The architecture of RobustScanner [13].

SAR: Figure 19 displays its architecture, combines a deep CNN for feature extraction with an RNN for sequence prediction, and a 2D attention mechanism that enables the model to focus on specific parts of the image, potentially leading to lower performance if it can't generalize well to the meter's text layout.

CRNN: Integrates a CNN for extracting visual features with an RNN for sequence modeling, making it

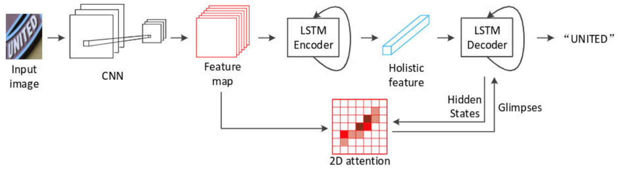


FIGURE 19. The architecture of SAR [14].

straightforward but potentially less adaptable to complex patterns and noise present in water meter images as illustrated in Figure 20.

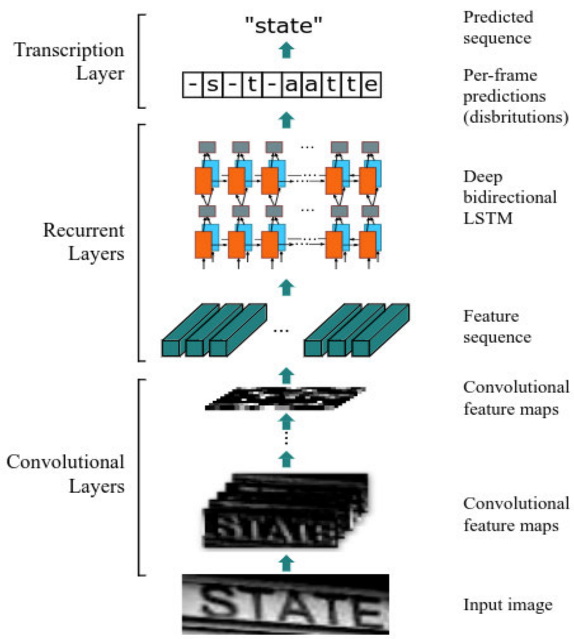


FIGURE 20. The structure of CRNN [36].

Models like PP-OCRv3 and TrOCR may outperform others due to their modern architectures that effectively capture complex dependencies and handle a variety of text presentations. The PP-OCRv3 benefits from a series of optimizations that contribute to its high accuracy, such as enhanced backbone networks and advanced training strategies which are specifically tailored to the task of OCR, providing it with a notable edge in performance. TrOCR's Transformer-based design, with its attention mechanisms, is particularly adept at handling complex patterns that require understanding both local and global contexts within an image, making it suitable for a diverse range of text recognition tasks. RobustScanner's approach to dynamically balance context and positional information can be advantageous in many scenarios but may not be as effective if the positional cues in the data are not pronounced or are inconsistent, as can be the case with some water meter readings. SPIN's focus on maintaining the structural integrity of the text may not be as critical for water meters where the text structure is relatively simple and consistent, possibly explaining its lower performance on this task compared to more versatile architectures. SAR's 2D attention mechanism allows it to focus on different parts of the image to decode characters but may struggle with non-standard text layouts or poor image quality, which are common in water meter readings. CRNN, being one of the earliest models integrating CNNs and RNNs, may not be as effective due to its relatively simpler design which doesn't include the more advanced attention mechanisms or architectural improvements of later models. Each OCR model was implemented on the cropped numeric regions with the following parameter settings as illustrated in Table 1.

TABLE 1. Each OCR model was implemented on the cropped numeric regions with the following parameter settings.

OCR Model	Framework	Parameters	Preprocessing
PP-OCRv3	PaddleOCR	Image size: 640×480, Batch size: 8, Learning rate: 1e-4, Epochs: 50	Histogram equalization, resizing
SPIN	PaddleOCR	Image size: 640×480, Learning rate: 1e-4, Epochs: 50	Resizing, Gaussian noise removal
TrOCR	PyTorch	Pretrained "base" model, Image size: 640×640, Learning rate: 5e-5, Epochs: 50	Adaptive thresholding
RobustScanner	MMOCR	Batch size: 8, Image size: 640×640, Learning rate: 1e-4, Epochs: 50	Resizing, contrast normalization
SAR	MMOCR	Batch size: 8, Image size: 640×480, Learning rate: 1e-4, Epochs: 50	Resizing
CRNN	PyTorch	Batch size: 16, Image size: 640×480, Learning rate: 1e-3, Epochs: 50	Resizing

Table 2 categorizes cases of digit recognition on water meters, from clear images with varying digit counts to images affected by shooting angles, blurriness due to water droplets, obstruction, unclear visibility, and half-digit transitions, aiding in the prediction of water consumption.

Table 3 provides the output of each OCR model for each case, comparing the performance, and showing how each model performs under various challenging cases for digit recognition, with correct outputs bolded for clarity.

Water meter readings with 4 to 8 digits are typically straightforward and easy to interpret. However, when encountering half-digit transitions (such as a number transitioning between 8 and 9), focusing on the lower portions of the digits becomes crucial for accuracy. Additionally, challenges arise from scenarios like unideal shooting angles, blurriness caused by water droplets, obstruction that partially covers the meter, unclear images, and ambiguous readings. These factors create significant hurdles, making it notably difficult to accurately interpret the displayed digits in such obstructed situations.

Environmental challenges such as poor lighting, water droplets causing blurriness, and obstructed digits significantly affect the accuracy of OCR models. In our study,

TABLE 2. Classification of issues in water meter recognition.

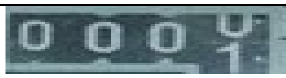
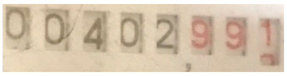
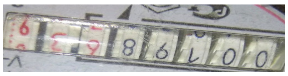
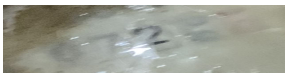
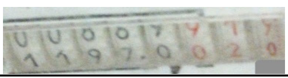
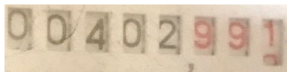
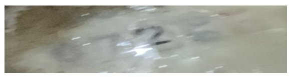
Cases	Images
4 digits	
5 digits	
6 digits	
7 digits	
8 digits	
Unideal shooting angle	
Blurriness Caused by Water Droplets	
Obstructed	
Unclear	
Half-digit transition	

TABLE 3. OCR model performance comparison in various conditions.

Cases	PP-OCRv3 TrOCR RobustScanner	SPIN SAR CRNN
	**0001** **0001** **0001**	0004 00011 0000
	**05609** **05609** **05609**	**05609** 056099 **05609**
	**079475** **079475** **079475**	017194715 **079475** **079475**
	**0009541** **0009541** **0009541**	0009540 **0009541** **0009541**
	**00402991** **00402991** **00402991**	**00402991** **00402991** **00402991**
	00588100 00168669 60984100	HEabDnon 008868 05810
	02 0022 001	1072E 0222 026
	00304555 00304565 00304155	00304850 0030455 00304555
	14024 14054406 14021944	TZUSAXK 402110 040286887
	**11970020** **11970020** **11970020**	HUOPEE 0086828 19999020

PP-OCRv3 demonstrated superior performance in low-light conditions due to its robust preprocessing modules like histogram equalization and adaptive thresholding. Similarly, **RobustScanner** effectively handled distorted text caused by angled shots by incorporating dynamic positional fusion. However, models like **SAR** and **CRNN** struggled with these scenarios due to their limited ability to adapt to non-ideal input images. Future iterations of this work could incorporate image enhancement techniques, such as deblurring algorithms and illumination correction, to mitigate these challenges and improve recognition accuracy.

The summary of considered framework can be analyzed as follows:

- Best Performance: PP-OCRv3, due to its efficient combination of feature extraction, attention mechanisms, and robust preprocessing capabilities.
- Scenario-Specific Models:
  - TrOCR: Best for long sequences.
  - RobustScanner: Best for low-light conditions.
- Limitations: SAR and CRNN were less effective due to their inability to adapt to challenging environmental conditions.

IV. CONCLUSION AND FUTURE WORK

This paper focuses on interpreting water meter readings through image analysis using deep learning techniques. The study demonstrates strong performance in accurately extracting and presenting readings from water meter images. However, challenges remain when handling blurry, dusty, or poorly captured images, which can affect reading accuracy. Addressing these limitations remains a critical challenge, requiring the exploration of alternative methods to enhance image recognition.

Looking ahead, the work aims to overcome these obstacles by developing more robust solutions for handling unclear images affected by dust, poor angles, or low quality. Future work will expand the system to interpret electrical meter readings, leveraging the existing deep learning model to handle various meter types. This solution can be practically deployed as a mobile application that allows end-users, such as utility workers or households, to capture images of water meters. The images can be processed in real-time or uploaded to a cloud server where the system integrates with existing billing or energy management systems. Additionally, incorporating the solution into IoT-enabled smart meters could further enhance automation, enabling seamless and frequent water consumption monitoring. Such a deployment would particularly benefit regions with low adoption of smart meters, as it can operate on traditional meters without infrastructure overhaul.

To ensure real-world utility and scalability, this system must address hardware and connectivity challenges. Deploying lightweight versions of the model optimized for mobile devices can extend its usability in resource-constrained environments. Additionally, incorporating offline processing capabilities allows the solution to operate effectively in rural or remote areas with limited internet connectivity. Generalizing the solution to handle diverse meter types across different regions will require further dataset expansion and fine-tuning on a wider variety of meter images.

Future work will focus on overcoming challenges posed by environmental factors, including poor lighting, low image resolution, and obstructions caused by water droplets or dirt. Integrating advanced image preprocessing techniques, such as deblurring algorithms, illumination normalization, and noise removal, can significantly enhance image quality. Additionally, the use of generative adversarial networks (GANs) for image enhancement could help improve recognition accuracy in adverse conditions.

ACKNOWLEDGMENT

(Bay Nguyen Van and Anh Nguyen contributed equally to this work.)

DISCLOSE STATEMENT

No potential conflict of interest was reported by the author(s).

REFERENCES

[1] L. Fiorini and M. Aiello, "Energy management for user's thermal and power needs: A survey," *Energy Rep.*, vol. 5, pp. 1048–1076, Nov. 2019, doi: 10.1016/j.egyr.2019.08.003.

[2] M. M. Salimuddin, D. K. Pandey, and B. K. Dubey, "Smart metering for smart power consumption," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 7, no. 3, pp. 1874–1878, Mar. 2019, doi: 10.22214/ijraset.2019.3350.

[3] O. Osanaiye, S. Unogwu, and F. Aina, "A GSM module-based smart electric meter reader," *Acta Electrotechnica Et Inform.*, vol. 20, no. 4, pp. 38–45, Jan. 2021, doi: 10.15546/aeei-2020-0024.

[4] D. Narciso-Celestino and N. I. Vargas-Cuentas, "Smart energy meter prototype for single-phase power lines in Peru," *Int. J. Eng. Trends Technol.*, vol. 69, no. 4, pp. 209–214, Apr. 2021, doi: 10.14445/22315381/ijett-v69i4p229.

[5] M. L. W. Concio, F. S. Bernardo, J. M. Opulencia, G. L. Ortiz, and J. R. I. Pedrasa, "Automated water meter reading through image recognition," in *Proc. IEEE Region Conf. (TENCON)*, Nov. 2022, pp. 1–6, doi: 10.1109/TENCON55691.2022.9977678.

[6] J. Yin, P. Xia, J. He, L. Gu, and K. Yang, "An automatic electricity meter reading system based on neural vision location," *J. Phys., Conf. Ser.*, vol. 1267, no. 1, Jul. 2019, Art. no. 012076, doi: 10.1088/1742-6596/1267/1/012076.

[7] M. Spichkova, J. van Zyl, S. Sachdev, A. Bhardwaj, and N. Desai, "Easy mobile meter reading for non-smart meters: Comparison of AWS rekognition and Google cloud vision approaches," in *Proc. 14th Int. Conf. Eval. Novel Approaches Softw. Eng.*, 2019, pp. 179–188, doi: 10.5220/0007762301790188.

[8] R. Haridas and R. L. Jyothi, "Convolutional neural networks: A comprehensive survey," *Int. J. Appl. Eng. Res.*, vol. 14, no. 3, p. 780, Feb. 2019, doi: 10.37622/ijaer/14.3.2019.780-789.

[9] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with YOLOv8," 2023, *arXiv:2305.09972*.

[10] C. Li, W. Liu, R. Guo, X. Yin, K. Jiang, Y. Du, Y. Du, L. Zhu, B. Lai, X. Hu, D. Yu, and Y. Ma, "PP-OCRv3: More attempts for the improvement of ultra lightweight OCR system," 2022, *arXiv:2206.03001*.

[11] C. Zhang, Y. Xu, Z. Cheng, S. Pu, Y. Niu, F. Wu, and F. Zou, "SPIN: Structure-preserving inner offset network for scene text recognition," 2020, *arXiv:2005.13117*.

[12] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, "TrOCR: Transformer-based optical character recognition with pre-trained models," 2021, *arXiv:2109.10282*.

[13] X. Yue, Z. Kuang, C. Lin, H. Sun, and W. Zhang, "RobustScanner: Dynamically enhancing positional clues for robust text recognition," 2020, *arXiv:2007.07542*.

[14] H. Li, P. Wang, C. Shen, and G. Zhang, "Show, attend and read: A simple and strong baseline for irregular text recognition," 2018, *arXiv:1811.00751*.

[15] Z. Zhao, "Research on Tibetan character recognition based on model CRNN," *Proc. SPIE*, vol. 12328, pp. 321–327, Sep. 2022, doi: 10.1117/12.2644284.

[16] G. Salomon, R. Laroca, and D. Menotti, "Deep learning for image-based automatic dial meter reading: Dataset and baselines," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8, doi: 10.1109/IJCNN48605.2020.9207318.

[17] G. Salomon, R. Laroca, and D. Menotti, "Image-based automatic dial meter reading in unconstrained scenarios," *Measurement*, vol. 204, Nov. 2022, Art. no. 112025, doi: 10.1016/j.measurement.2022.112025.

[18] F. Martinelli, F. Mercaldo, and A. Santone, "Water meter reading for smart grid monitoring," *Sensors*, vol. 23, no. 1, p. 75, Dec. 2022, doi: 10.3390/s23010075.

[19] C. Li, Y. Su, R. Yuan, D. Chu, and J. Zhu, "Light-weight spliced convolution network-based automatic water meter reading in smart city," *IEEE Access*, vol. 7, pp. 174359–174367, 2019, doi: 10.1109/ACCESS.2019.2956556.

[20] M. Imran, H. Anwar, M. Tufail, A. Khan, M. Khan, and D. A. Ramli, "Image-based automatic energy meter reading using deep learning," *Comput., Mater. Continua*, vol. 74, no. 1, pp. 203–216, 2023, doi: 10.32604/cmc.2023.029834.

[21] Q. Hong, Y. Ding, J. Lin, M. Wang, Q. Wei, X. Wang, and M. Zeng, "Image-based automatic watermeter reading under challenging environments," *Sensors*, vol. 21, no. 2, p. 434, Jan. 2021, doi: 10.3390/s21020434.

[22] R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz, and D. Menotti, "Convolutional neural networks for automatic meter reading," *J. Electron. Imag.*, vol. 28, no. 1, p. 1, Feb. 2019, doi: 10.1117/1.jei.28.1.013023.

[23] Y. Liang, Y. Liao, S. Li, W. Wu, T. Qiu, and W. Zhang, "Research on water meter reading recognition based on deep learning," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 12861, doi: 10.1038/s41598-022-17255-3.

[24] K. Koščević and M. Subašić, "Automatic visual reading of meters using deep learning," in *Proc. Croatian Comput. Vis. Workshop*, Feb. 2019, pp. 1–6, doi: 10.20532/ccvw.2018.0002.

[25] H. Liu, Y. Li, and D. Liu, "Object detection and recognition system based on computer vision analysis," *J. Phys., Conf. Ser.*, vol. 1976, no. 1, Jul. 2021, Art. no. 012024, doi: 10.1088/1742-6596/1976/1/012024.

[26] B. Kaur and S. Singh, "Object detection using deep learning," in *Proc. Int. Conf. Data Sci., Mach. Learn. Artif. Intell.*, Aug. 2021, pp. 328–334, doi: 10.1145/3484824.3484889.

[27] S. Faizullah, M. S. Ayub, S. Hussain, and M. A. Khan, "A survey of OCR in Arabic language: Applications, techniques, and challenges," *Appl. Sci.*, vol. 13, no. 7, p. 4584, Apr. 2023, doi: 10.3390/app13074584.

[28] R. U. Khan, X. Zhang, and R. Kumar, "Analysis of ResNet and GoogleNet models for malware detection," *J. Comput. Virol. Hacking Techn.*, vol. 15, no. 1, pp. 29–37, Mar. 2019, doi: 10.1007/s11416-018-0324-z.

[29] Ultralytics. (Dec. 27, 2023). *GitHub—Ultralytics/ultralytics: NEW—YOLOv8 in PyTorch > ONNX > OpenVINO > CoreML > TFLite*. [Online]. Available: https://github.com/ultralytics/ultralytics

[30] *COCO—Common Objects in Context*. Accessed: Mar. 15, 2024. [Online]. Available: https://cocodataset.org/#home

[31] Y. Du, Z. Chen, C. Jia, X. Yin, T. Zheng, C. Li, Y. Du, and Y.-G. Jiang, ''SVTR: Scene text recognition with a single visual model,'' 2022, *arXiv:2205.00159*.

[32] C. Cui, T. Gao, S. Wei, Y. Du, R. Guo, S. Dong, B. Lu, Y. Zhou, X. Lv, Q. Liu, X. Hu, D. Yu, and Y. Ma, ''PP-LCNet: A lightweight CPU convolutional neural network,'' 2021, *arXiv:2109.15099*.

[33] W. Hu, X. Cai, J. Hou, S. Yi, and Z. Lin, ''GTC: Guided training of CTC towards efficient and accurate scene text recognition,'' 2020, *arXiv:2002.01276*.

[34] X. Zhang, B. Zhu, X. Yao, Q. Sun, R. Li, and B. Yu, ''Context-based contrastive learning for scene text recognition,'' in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 3, pp. 3353–3361, doi: 10.1609/aaai.v36i3.20245.

[35] J. Baek, Y. Matsui, and K. Aizawa, ''What if we only use real datasets for scene text recognition? Toward scene text recognition with fewer labels,'' 2021, *arXiv:2103.04400*.

[36] LearnOpenCV-Learn OpenCV, PyTorch, Keras, Tensorflow With Examples Tutorial. (Jun. 14, 2022). *Optical Character Recognition Using PaddleOCR | LearnOpenVC*. [Online]. Available: https://learnopencv.com/optical-character-recognition-using-paddleocr/

BAY NGUYEN VAN is currently a Lecturer with Ho Chi Minh City Open University. His research interests include machine learning and OCR.

ANH NGUYEN received the master's degree from the FPT University, in 2024. His research interests include security, machine learning, and pattern cognition.

KIET TRAN-TRUNG received the master's degree from Ho Chi Minh City Pedagogical University, Vietnam. He is currently a Lecturer with Ho Chi Minh City Open University. His research interests include machine learning and computer vision.

THIEN HO HUONG is currently a Lecturer with Ho Chi Minh City Open University. His research interests include machine learning and text mining.

HA DUONG THI HONG is currently a Lecturer with Ho Chi Minh City Open University. Her research interests include machine learning and feature selection.

HAU NGUYEN TRUNG received the master's degree from the University of Science, Vietnam. He is currently a Lecturer with Ho Chi Minh City Open University. His research interests include machine learning and computer vision.

VINH TRUONG HOANG received the master's degree from the University of Montpellier, in 2009, and the Ph.D. degree in computer science from the University of the Littoral Opal Coast, France. He is currently an Assistant Professor and the Head of the Image Processing and Computer Graphics Department, Ho Chi Minh City Open University, Vietnam. His research interests include image analysis and feature selection.