Deep Learning Based Automated Electronic Meter Reading System using YOLOv5 Architecture

Harsh Gupta
Ankita Jaiswal
Pavinder Yadav
20mma007@nith.ac.in
193048@nith.ac.in
pavinder_phdmath@nith.ac.in
National Institute of Technology Hamirpur
Hamirpur, Himachal Pradesh, India

Nidhi Gupta*
National Institute of Technology Kurukshetra
Kurukshetra, Haryana, India

ABSTRACT

Automatic instrument reading has become a critical issue for intelligent sensors in smart cities. Several artificial intelligence techniques are developing tools for addressing the issue. The image-based Automatic Meter Reading (AMR) techniques have been tested on images taken under regulated conditions, but they become unresponsive when dealing with fuzzy, hazy or blurred meter images. In this paper, we deal with AMR, which focuses on unconstrained settings such as fuzzy, hazy or blurry meter images. Automated meter reading consists of three major components: identifying the counter region, localising and cropping the counter region and digit recognition. In this article, the deep learning model YOLOv5 have been used on the image dataset. YOLOv5 is a state-of-the-art single-stage deep learning detector that outperforms all other detectors and it is observed that the proposed technique and the trained model based on YOLOv5 can reliably detect and recognise meter readings from the different meter kinds. For the task of digit recognition, a YOLOv5 based custom-built digit optical character reader is used that can recognise 0-to-9-digit numbers. Furthermore, the proposed AMR system achieves remarkable recognition rates of 99.74% for counters and 88.70% for digit recognition even while rejecting counters with lower confidence values.

CCS CONCEPTS

• **Computing methodologies** → *Neural networks.*

KEYWORDS

Deep Learning, Automatic Meter Reading, Object Detection, Automated Detection, Convolution Neural Network, Image Segmentation

*Corresponding Author.

BDMIP '23, November 17–19, 2023, Xiamen City, China, China

ACM ISBN 979-8-4007-0916-6/23/11
https://doi.org/10.1145/3645279.3645288

ACM Reference Format:
Harsh Gupta, Ankita Jaiswal, Pavinder Yadav, and Nidhi Gupta. 2023. Deep Learning Based Automated Electronic Meter Reading System using YOLOv5 Architecture. In *International Conference on Big Data Mining and Information Processing (BDMIP '23), November 17–19, 2023, Xiamen City, China, China.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3645279.3645288

1 INTRODUCTION

Automatic Meter Reading (AMR) is a technology that makes it possible for water meters or energy meters to automatically detect and record usage, diagnostic and status data in a central database for troubleshooting, billing, and further important analysis. This method primarily helps utility companies to save financial resources by removing the requirement to make frequent trips to each physical site in order to read meters with cost-efficient parameters. Another significant advantage is that billing may be done on the basis of consumption in close proximity to the moment it occurs rather than on predictions based on usage in the past or in the future. When paired with analysis, this real-time data can assist utility companies and their customers in more successfully managing the use of electric energy, gas, water, etc. The concept of image-based AMR is a specialised instance of the scene text detection and identification problem, which is predicated on the assumption that the inspection can be carried out automatically with the goal of reducing the number of mistakes made by humans and minimising the amount of required human intervention.

Despite the fact that image-based AMR has received a great deal of attention in the recent years. Most of the studies in the literature are still found limited in a variety of ways. The research was conducted using either private datasets or datasets including images shot under controlled settings. In contrast, recent related research on automatic number plate recognition has shifted the focus to unconstrained scenarios, significantly improving the state-of-the-art. These scenarios have problems like blurring, different lighting conditions, varying sizes, in-plane and out-of-plane rotations, occlusions, and so on. Additionally, the majority of studies focus on just one step of the AMR pipeline, making it difficult to conduct a comprehensive evaluation of the approaches that are available from the beginning to the extreme end. For example, dependent on the how accurately the counters area are detected, the results of a recognition model may vary substantially.

The objective of this research is to use the YOLOv5 architecture to create and deploy an automated electronic meter reading

(AMR) system based on deep learning. The main goal is to develop a high-performing AMR system with three main components: first, identifying the counter region in the meter images; second, cropping the region of interest to enable digit recognition; and third, using a customized OCR method based on the YOLOv5 deep learning model to identify individual digits.

In the past several years, Deep Convolutional Neural Networks (DCNNs), or deep learning techniques, have surpassed all conventional machine learning techniques in object identification. Finding higher level characteristics in data requires a lot of effort using conventional approaches [8]. DCNN models accomplish this task automatically. These methods are used in a variety of fields, including self-driving auto-mobiles [9], traffic monitoring [10], weapons recognition systems [28, 29], face recognition systems [11], natural language processing [30], license plate detection [6] and many others.

In this work, an AMR system is developed, which consists of three stages. In order to achieve ultimate objectives, the counter region is detected, then the region of interest is cropped. The customised Optical Character Recognition (OCR) based on the deep learning model YOLOv5 is used on the new dataset made from the images of meters that have been cropped. YOLOv5 have been used to detect the individual digits. It is an object recognition system that identifies numerous objects in a single frame in real time. The trained model identifies things more accurately and quickly than previous existing architectures. It can identify up to 9,000 classes, including previously unidentified digits as well. The proposed scheme is summarised as follows:

- Manually collected and annotated 7,877 images of meters for counter detection.
- YOLOv5 is utilised for counter detection as well as recognition in a two-stage AMR system.
- A custom function is utilised for the detection of counters and cropping of the counter region, thus making a new dataset of cropped meter images.
- This new dataset was again annotated for each digit and trained using Yolov5 based custom OCR for digit recognition.

The organization of the paper consists of the following sections. Section 2 focuses on the related work of automatic meter reading. Section 3 describes the datasets and materials used in this research work. Section 4 describes the deep learning algorithms employed in this research work. Further, Section 5 describes the proposed methodology in detail. The results analysis of outcomes and discussion of deep learning architectures are discussed in Section 6. In last, conclusion along with future scope is given in Section 7.

2 RELATED WORK

Over the last decade, various methods for automatic meter reading using meter images have been proposed. Several strategies were utilised in the previously proposed methodologies, such as augmentation, contouring, segmentation, and so on, but the results were not as effective as compared to using deep learning architectures.

AMR application relies on text detection and recognition from meter images. However, the following essential changes should be highlighted in order to locate and identify the entire scene text: (i) Each AMR image has only one area of interest, known as the counter region, which contains all the numbers. (ii) AMR recognition networks must learn all digit classes (digits 0-9), whereas general-scene text recognition networks must learn both letters and digits. AMR presents a distinct problem for OCR systems. Even when sophisticated digit identification algorithms are used, rotated digits are typically the source of errors in such meters.

The importance of datasets is quite clear. Under various conditions, strong deep learning techniques that have only been trained on large datasets of images (such as ICDAR 2013) [13] are likely to fail. In this section, a description of the related works are provided that have been carefully looked at. The task of localization and detection of Racing Bib Numbers (RBN) has been performed by Wong *et al.* [27] in two phases. The first phase task of localization was achieved using YOLOv3 and RBN text recognised using a Convolutional Recurrent Neural Network (CRNN). The used model achieved a mean average precision (mAP) of 85.15%, however, in the unconstrained conditions the model failed to recognise RBN.

A deep learning model was employed by Son *et al.* [23] for number detection and recognition for Gas Meter Reading (GMR). This system was employed for both detecting meter reading and ID region for meters installed in Korea. The dataset consisted of 5,000 images of gas meters and obtained an accuracy of 85.71% for the counter region and 60.90% for the ID region.

Laroca *et al.* [17] created a two-stage AMR system in 2019 that utilised CNNs for both counter detection and recognition. The proposed UFPR-AMR dataset consists of 10,000 digits in total from 2,000 images of digits, each containing five digits, that have been annotated manually. The dataset also has a clear way to evaluate different approaches. The model achieved an accuracy of 94.13% with a reasonable frame rate (fps) only on high-end graphics cards.

Again, Laroca *et al.* [16] continued their work and modified the state-of-the-art AMR system and introduced a new technique called corner detection and classification for detecting the corners of the counter. The approach further increased the accuracy of the UFPR-AMR and Copel-AMR datasets (consisting of 12,500 manually annotated meter images in total) to 94.75% and 96.98%, respectively. However, the model still observed errors in the case of rotated or flipped images. For counter recognition, Li *et al.* [18] put forward a lightweight CNN that combines a fixed number of 1x1 and 3x3 kernels to minimise the parameters of the network while maintaining a low loss in recognition rate, given the requirement to design extremely efficient algorithms in the AMR environment. The experiments, however, were conducted solely on a private dataset with well-controlled photos that were almost similar to one another, i.e., the images were collected by a camera installed in the meter box and manually pre-processed by the authors. Pre-processing includes blurring, changing sizes, shadows and occlusions in the datasets.

Marques *et al.* [4] optimised the Faster R-CNN [22] and the RetinaNet [19] object detectors so that they could perform better in the counter recognition task. In the experiment, a tiny subset of counter pictures was taken from a private dataset. The mAP rates for both detectors are observed to be over 90%.

Koscevic and Subasic [15] made use of Faster R-CNN in order to recognise counters and serial numbers on the images of residential meters. Due to the fact that the dataset was compiled manually and comprised of two categories, gas and electricity, this was one of

the first research to be claimed on the GMR. The dataset included a total of 312 images from the meter class and 463 images from the power category. The model was able to correctly read 97.01% of the counter region and 63.01% of the serial number region.

On the other hand, Tsai *et al.* [24] employed a fine-tuned Single Shot MultiBox Detector (SSD) [21] to detect counters in the power meters. The model was able to get a score of 100% on the dial detection stage using both YOLO and Faster R-CNN. The identification rates for digits were 93.60%, while the recognition rate for meters was 75.25%.

YOLOv5 was used by Kasper *et al.* [14] to identify heavy freight trucks at winter rest places to estimate parking spot occupancy. YOLOv5 was also put into action by Benjumea *et al.* [3] for detecting small objects for autonomous vehicles, which has a maximum accuracy of 96.05% by improving YOLOv5 and is called YOLO-Z.

3 DATASET AND MATERIAL

In the presented work, the AMR dataset consists of over 8,000 electronic meter images collected by an energy company located in India. The meter images are collected from various Indian households, including districts, villages and settlements. As a result, AMR is made up of images taken in unrestricted situations, which generally contain blurriness (due to camera movement), ambiguity, variable sizes, in-plane and out-of-plane rotations, reflections, shadows and occlusions as shown in Figure 1. Because of the obstructions or broken meters, the meter reading could not be successfully completed on 800-1000 images. The original dataset consisted of several hazy images as well, which were eliminated manually, bringing the total number of images in the cleaned dataset of 7,877 images. The curated dataset used in this research included 7,877 images in total after the data was cleaned. This dataset was partitioned using an approach that assigned 6,300 images for training and the remaining 1,577 images for validation. The images have resolutions of 3072 x 4096. All the images are manually labelled over all the meter readings to mark the position of each of the four corners of the counter and a bounding box of dimension (x, y, w, h) for the counter region is developed. The second dataset, made up of 1,000 images of cropped counters, is generated from the original dataset by employing a custom function, which crops the counter region and saves the cropped images automatically.

Figure 1: Sample images of AMR dataset.

Since in the new dataset, each meter consists of 7 digits, a total of 5,537 digits were labelled manually. It has been observed that the digit '0' appears more frequently than the other digits, which is to be expected given that a brand new meter starts with '0000000' and it takes longer to increase the leftmost digit position. As a result, the frequency of numbers decreases with increasing numbers (i.e., digit 0 has the highest frequency and digit 9 has the lowest).

4 DEEP LEARNING ARCHITECTURE

Deep learning is an artificial neural network-based branch of machine learning science [2]. Multi-Layer Perceptrons (MLP), CNNs, and Recurrent Neural Networks (RNN) are some of its variants that can be used in many applications, such as natural language processing, computer vision based applications, and machine translation. The applications of deep learning has evolved gradually over the last decade in these areas. Neural networks had to evolve structurally from prior network designs, along with a lot more processing power. Some of the aspects of neural network evolution are: (i) The current networks include a greater number of neurons than the previous networks; (ii) There are more advanced methods for interconnecting neural network layers and neurons; (iii) Massive increase in available processing power for training; and (iv) Automatic feature extraction in deep learning. The architecture of the simple neural network is shown in Figure 2. An input layer, a hidden layer, and an output layer are the components that make up this structure. The initial data of the neural network has been stored in the input layers. The hidden layers are the layers between the input and output layers that does all of the important computational tasks, hence it plays the most important role in producing a result for given inputs in the output layer.

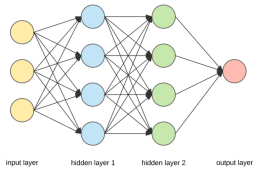


Figure 2: Sample neural network. [1]

The goal of the algorithms used in deep learning is to mimic how the human brain learns and processes information. The term "deep learning" refers to neural networks that are organised into one of four fundamental network designs and include a significant number of parameters and layers.

- Pre-trained unsupervised networks
- Neural networks with convolutions
- Neural networks with recurrent connections
- Neural networks that recurse

It is widely known that deep learning is a subset of machine learning. In reality, machine learning and deep learning work in a similar way. A significant advantage of deep learning over typical machine learning methods is automatic feature extraction. Basic machine learning models get better at executing their respective tasks when new data is provided. But they still require human involvement and interaction. Machine learning practitioners have traditionally spent months, years and even decades manually constructing exhaustive feature sets for data classification. With the intervention and required changes, an incorrect forecast could be improved. But an algorithms of deep learning perform the decision and prediction automatically without human error possibilities. This process is better explained in Figure 3, where feature extraction and classification are done by the network itself.

Deep learning has outperformed traditional algorithms in terms of accuracy and consistency in every type of data set with minimal

human effort and adjustments. These deep networks can assist data science teams by saving time and effort for more important tasks. CNNs have been utilised in this work because of their ability to create a two-dimensional internal representation of an image.

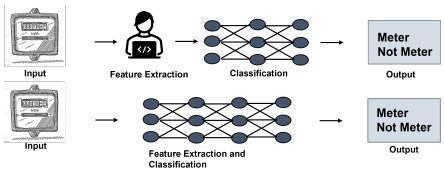


Figure 3: Difference between machine learning and deep learning.

4.1 Convolutional Neural Networks

CNNs are types of deep learning model used to analyse information with a grid pattern like images. They are based on the biological structure of visual cortex of humans and are meant to automatically and adaptively learn the spatial hierarchies of characteristics from low to high level patterns [7]. In linear algebra, specifically in matrix multiplication, CNNs are used to find the patterns in the images. This makes the process easier to use for image classification and object identification.

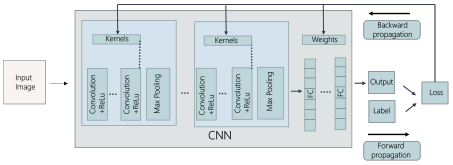


Figure 4: CNN architecture.

The building pieces that make up a CNN are detailed in the CNN architecture, as shown in Figure 4. These building blocks include convolution layers, pooling layers (such as max pooling) and fully connected layers. On a training dataset, a loss function and forward propagation are used to figure out how well the proposed model with certain kernels and weights works. Meanwhile, the performance of the proposed model with learnable parameters, such as kernels and weights, is computed by employing a gradient descent optimization procedure.

4.2 YOLOv5

YOLOv5 [12] is enhanced version of YOLOv4. The algorithm has been continually improved, because of this, it has received the highest scores possible on two official object recognition datasets: Microsoft Common Objects in Context (Microsoft COCO) [20], as well as Pascal Visual Object Classes (Pascal VOC) [5]. YOLOv5 is distinct from its predecessors since it is based on PyTorch rather than being a fork of the initial Darknet. Similar to the YOLOv4, the YOLOv5 is outfitted with both a CSP framework as well as a PANet head. The most significant enhancements are auto-learning bounding box anchoring and mosaic data augmentation.

Figure 5 illustrates the YOLOv5 network. It is comprised of three components: the backbone (CSPDarknet), the neck (PANet), and the head (YOLO Layer). Before being sent to PANet for feature fusion, initially, the data is sent over CSPDarknet to have its features extracted.

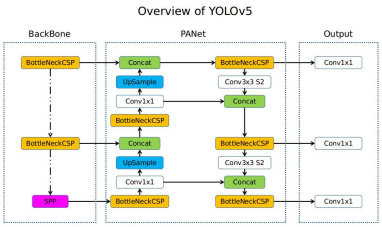


Figure 5: The network architecture of YOLOv5.

The YOLO Layer will then send out the results of its detection: the object's category, class score, location, and bounding box size. CSPNet [25] was merged into Darknet through YOLOv5, resulting in CSPDarknet as its backbone. By integrating the gradient alterations to the feature space and resolving the issues with the repeated gradient information in large-scale backbones, CSPNet decreases model parameters and floating-point computations per millisecond, assuring precision and speed while lowering model size.

In the digit identification job, detection accuracy and speed are crucial and compact model size also affects how well it can draw conclusions on edge devices with limited resources. To increase the information flow, the YOLOv5 used PANet [26] as its neck. The usage of a unique feature pyramid network (FPN) architecture by PANet with an enhanced bottom-up route increases the transmission of low-level characteristics.

The choice of anchor boxes in YOLOv5 for object detection is driven by an innovative approach that dynamically computes anchor box dimensions using a K-means clustering algorithm during training. Unlike previous versions, this adaptive method identifies anchor box sizes best suited to represent the varied object scales and aspect ratios present within the dataset. This automated process eliminates the need for manual selection or tuning of anchor boxes, enhancing the model's adaptability to diverse object sizes and shapes while improving object localization accuracy and overall detection performance.

The feature grids and all feature levels are connected via the use of adaptive feature pooling, which enables the direct transfer of essential information through one feature level to the next sub - network. The usage of accurate localization signals is improved in lower layers by PANet, which may clearly improve object location accuracy. The model can handle small, medium, and large objects because it comes with three different sizes of feature maps (18, 36, and 72). This is part of the multi-scale features.

5 PROPOSED METHODOLOGY

This section includes the detailed methodology of the proposed work of AMR. It is divided into several major steps; dataset collection and the method used for the detection and recognition of

digits. Overall, from beginning to end, the input image is taken, annotated for the counter region and then applied to YOLOv5. The images are cropped by the counter using the obtained coordinates. The cropped counter is annotated with each digit to recognise the meter reading. The output meter reading is then obtained by using the YOLO architecture. The basic workflow of the proposed scheme is depicted in Figure 6.

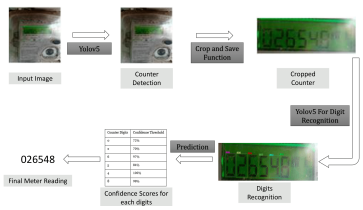


Figure 6: An end-to-end approach applied for Automatic Meter Reading.

As a critical preparatory measure preceding the commencement of model training, all images within the dataset underwent meticulous resizing to a standardized dimension of 416 x 416 pixels. This uniformity in image size is pivotal for optimizing the subsequent training of the YOLOv5 model for object detection tasks. Following this resizing process, an essential normalization step was applied, ensuring that the pixel values across all images were scaled to a consistent range between 0 and 1. This normalization strategy holds paramount importance as it significantly aids in expediting the convergence process during the model's training phase. By standardizing the pixel values within this designated scale, the model can learn more efficiently, discern intricate visual patterns, and extract pertinent features from the dataset. These deliberate preprocessing steps are instrumental in ensuring uniformity, enhancing computational efficiency, and fortifying the model's capacity to adeptly detect objects across diverse contexts, thereby contributing significantly to its overall robustness and performance.

YOLOv5 uses a cloud-based data augmentation approach called a data loader to handle training data in batches during training. Three different augmentation techniques are used by this loader: scaling, color space modifications, and mosaic augmentation. The most creative of these techniques is mosaic augmentation, which mixes four images in a complicated way by splitting them into four tiles with variable aspect ratios. This enriches the dataset by synthesizing various contextual information within each batch.

The proposed approach is divided into three main parts: i.) Counter detection, ii.) Localization of counter region and cropping iii.) Detection and recognition of digits. The YOLOv5 model is utilised to determine the counter region in given input image,. After that, we crop the counter using the coordinates of detector. Finally, counters are delivered to the Yolov5-based custom-built OCR recognition network using the coordinates of detector .

5.1 Counter Detection

In various illumination conditions, it is difficult to locate the counter immediately. Some text blocks, like meter specifications and serial numbers, look a lot like the counter region in some meter models. YOLOv5 model is employed in an image to determine the meter reading zone. The annotated set of 7,877 images is further partitioned into training and testing datasets for the model. The training set contains 6,302 images, while the testing set has 1,575 images. The training process takes around 10 hours to complete and the best-weighted file accuracy attained is 99.78% for all sorts of images including complex ones that are blurry and with unclear visibility.

5.2 Localization of Meter Region and Cropping

The best weight file has been deployed, which produced by YOLOv5 after training to determine the meter reading region of all images. The obtained coordinates of the detected corner points of region are then utilised to crop the counter region from the image. Consequently, another set of cropped image datasets is constructed in this stage, which are further utilised to train a new digit detection and recognition model later.

5.3 Digit Detection and Recognition

Cropped images of meter readings are used for digit detection. Another set of labelled datasets is constructed for digit detection with 1,000 cropped images marked 0–9. The digit labelled dataset is then trained using YOLOv5 architectures based custom OCR. The training set images include the cropped images of meter digits which are hard to see, due to several constraints.

6 RESULTS AND DISCUSSION

The proposed model is divided into major three parts: counter detection, localization and cropping, followed by the digit recognition. The experimental performance of the proposed AMR system is discussed in detail in this section. The training and testing data images for the work are taken from various meter devices (analogue and digital) and consist of total of 7,877 images. Initially, the counter region test is started in which mAP is calculated to determine how accurately the counter region is being recognised, where mAP is a metric used to evaluate object detection models such as Fast R-CNN, YOLO, Mask R-CNN and others. The mAP values are derived using recall values ranging from 0 to 1. There is a potential that the model detects the counter incorrectly sometimes, hence true positive and false positive values are also calculated to examine the detection accuracy of the model with right bounding box region.

The YOLOv5 framework is used in an Automated Material Recognition (AMR) system, where it facilitates both counter detection and recognition tasks in a two-stage procedure. The model focuses on recognizing a single class in the first stage, known as counter detection. Following that, in the second stage, the system is tasked with recognizing and localizing digits spanning from 0 to 9, consequently encompassing 10 separate detection and classification classes.

Precision, recall and F1 score are metric scores derived from the confusion metric, where precision is a measure of how many true positive predictions are produced and recall is a measure of how many positive instances the classifier properly predicted out of all the positive detections from the model. The F1 score is a metric that combines accuracy and recall. It is the arithmetic mean of accuracy and recall. The Intersection over Union (IoU) method is also used to compute the mAP parameter. It ranges between 0 and

1, which reflects the overlapping between the predicted and actual bounding boxes. The premitive performance parameters namely precision, recall, F1-score and accuracy are measured by Eq. 1 to Eq. 4, respectively.

$$precision = \frac{TP}{TP + FP} \tag{1}$$

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1\_score = 2 * \frac{precision * recall}{precision + recall} \tag{3}$$

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{4}$$

where, TP, FP, TN, FN stands for True Positive, False Positive, True Negative and False Negative, respectively. Figure 7 shows an example of a processed image after the counter detection. It is visible that blurred and tilted meter images are also being processed successfully in the experiment. Figure 8 depicts some of the final cropped images from the dataset. This dataset contains only the region of digits to be further detected and identified by the model.

Figure 7: Counter detection sample images.

Figure 8: Sample cropped images.

The testing dataset contained 290 images, which is divided into 32 batches and 200 epochs. The results of YOLOv5 are shown below in the Table 1. Further, the mAP is calculated at a threshold value of 0.5.

The robustness of the counter detection model is critical for the efficient performance of the digit detection and recognition models.

Table 1: YOLOv5 Performance

ClassID	Labels	Precision	Recall	F1-Score	mAP@.5
All	1643	0.898	0.828	0.862	0.887
0	323	0.822	0.884	0.852	0.875
1	191	0.888	0.812	0.849	0.877
2	183	0.829	0.765	0.796	0.854
3	130	0.931	0.815	0.871	0.904
4	148	0.977	0.863	0.918	0.914
5	141	0.856	0.773	0.813	0.85
6	157	0.951	0.871	0.910	0.937
7	104	0.924	0.837	0.879	0.875
8	138	0.923	0.848	0.884	0.91
9	128	0.878	0.812	0.844	0.87

The counter detection model with images in an unconstrained scenario yield the highest mAP of 99.78%. The model took an average of 26 miliseconds to detect the counters in an image. The prediction process of the cropped image digits and confidence score detected digits from the image is mentioned in Figure 9.

```
Detection layer: 82 - type = 28
Detection layer: 94 - type = 28
Detection layer: 106 - type = 28
/content/drive/MyDrive/yolodigit/1.png: Predicted in 26.666000 milli-seconds.
0: 87%
0: 81%
4: 99%
0: 90%
5: 98%
4: 99%
0: 58%
```

Figure 9: Output confidence value

Figure 10 shows several sample output images of digit detection from challenging input images. As the digits are not visible in all the cropped images, detection and recognition is a difficult task and hence infeasible.

Figure 10: Sample digit prediction.

The testing image dataset has an overall label count of 1,643 images with a combined precision as 89.80%, recall as 82.80% and an mAP as 88.70%. Figure 11 represents the confusion matrix of all the digits giving the impression of out-performance of training on YOLOv5. The true prediction of digits '0' and '4' surpasses the true prediction of all other digits. The background False Negative (FN) row and background False Positive (FP) column give the information about incorrect predictions made by the detector, which is low for all the digits, giving an impression of accuracy of the proposed detector.

Figure 12 highlights the F1 score of all the individuals as well as of all the classes. The detector has an overall F1 score of 86%

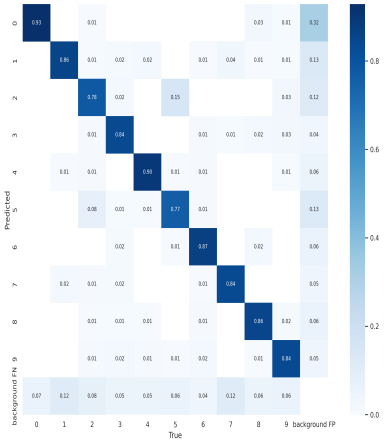


Figure 11: Confusion matrix.

at 57.30% confidence value. Figure 13 represents the trade-off between precision and recall for different threshold values. Since the area between the Precision-Recall (PR) curves for all the classes is around 0.887, it represents both high precision and high recall value. High scores for both show that the detector is performing well and returning accurate results (high precision), as well as high recall value. It can also be observed from the curves that the digit '6' has the maximum F1 score and PR curve area, while the digit '2' has the minimum F1 score and PR area.

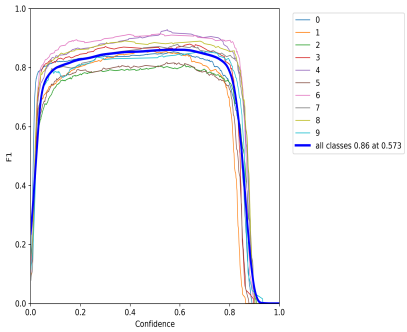


Figure 12: F1 curve.

The performance metric of both the training and validation sets is given in Figure 14. Basically, 3 kind of losses exist, namely, box loss, object loss and classification loss. The model accuracy is defined by locating the centre of an object as well as prediction of bounding boxes that covers the object. Another loss is based on the measurement in terms of probability of existence of an object in the proposed region. The high value defines the likelihood of object in the image window and vice versa. Another classification loss is defined by prediction of accuracy of belongingness to the particular class. these losses observed rapid decline until around 150 to 200 epochs. The sample images of training and validation by YOLOv5 are shown in Figure 15 and Figure 16, respectively.

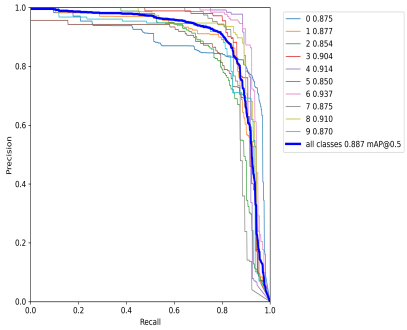
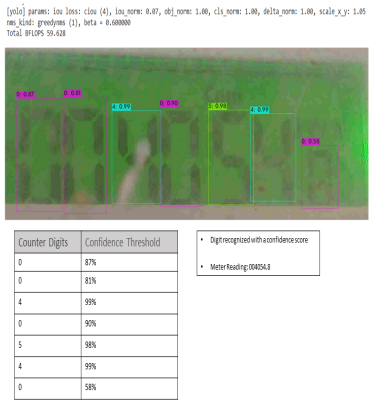


Figure 13: Precision recall curve.



Counter Digits	Confidence Threshold
0	87%
0	81%
4	99%
0	90%
5	98%
4	99%
0	58%

Figure 14: Plots of loss on training and validation set.

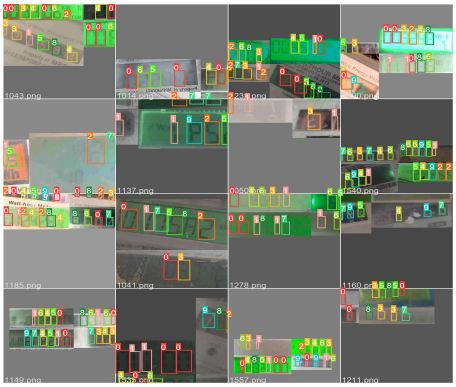


Figure 15: Sample training images.

7 CONCLUSION AND FUTURE SCOPE

In this research work, an efficient strategy for an automated meter reading system is presented that yields state-of-the-art counter detection results as well as consistent results for various illumination and unconstrained images. The major contribution of this research

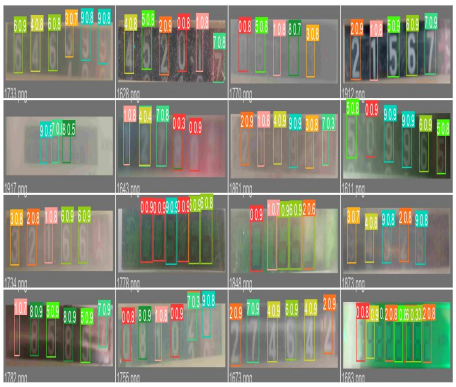


Figure 16: Sample validation images.

is the development of a systematic approach to automating the meter reading, which was accomplished using a step-by-step approach for which the training dataset from a set of 8,000 meter reading images containing diverse unconstrained scenarios are used.YOLOv5 is used to train the meter images for the counter detection and then use the weights for counter detection and detecting the meter reading region to generate cropped images of the meter reading. The accuracy of the trained model was determined to be 99.74%. The cropped image after digit recognition has an excellent accuracy 88.70% in YOLOv5.

Deep Learning-Based Automated Electronic Meter Reading Systems using YOLOv5 architecture include limited evaluation across diverse meter types and conditions, a lack of optimization for large-scale deployment, unexplored adaptation to varied data augmentation techniques, and insufficient investigation into contextual integration and robustness against adverse environmental factors.

The future work plan is to build a technique for detecting the counter region and its corners at the same time to improve the speed as well as the accuracy. The AMR pipeline can be used to investigate the performance of automated recognition meter models and types as well as cases in which the counter has rotated digits, which is a common cause of reading errors in the analogue meters. Secondly, the dataset size needs to be increased to improve the accuracy.

FUNDING INFORMATION

This research did not receive any specific grants from funding agencies in the public, commercial or non-profit sectors.

CONFLICT OF INTEREST

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

DATA AVAILABILITY STATEMENT

The datasets analysed during the current study are not publicly available due to licencing under government organisation but are available from the corresponding author on reasonable request.

REFERENCES

[1] 2018. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 4, 11 (2018), e00938.

[2] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data* 8, 1 (2021), 1–74. https://doi.org/10.1186/s40537-021-00444-8

[3] Aduen Benjumea, Izzedin Teeti, Fabio Cuzzolin, and Andrew Bradley. 2021. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. *arXiv preprint arXiv:2112.11798* (2021).

[4] Ricardo Costa da Silva Marques, Arthur Costa Serra, João Vitor Ferreira França, João Otavio Bandeira Diniz, Geraldo Braz, João Dallyson Sousa de Almeida, Marcia Izabel Alves da Silva, and Eliana Marcia Garros Monteiro. 2019. Image-based electric consumption recognition via multi-task learning. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. 419–424. https://doi.org/10.1109/BRACIS.2019.00080

[5] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88, 2 (2010), 303–338. https://doi.org/10.1007/s11263-009-0275-4

[6] V Gnanaprakash, N Kanthimathi, and N Saranya. 2021. Automatic number plate recognition using deep learning. In *IOP Conference Series: Materials Science and Engineering*, Vol. 1084. IOP Publishing, 012027.

[7] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. 2018. Recent advances in convolutional neural networks. *Pattern Recognition* 77 (2018), 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

[8] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew. 2016. Deep learning for visual understanding: A review. *Neurocomputing* 187 (2016), 27–48.

[9] Abhishek Gupta, Alagan Anpalagan, Ling Guan, and Ahmed Shaharyar Khwaja. 2021. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array* 10 (2021), 100057.

[10] Himanshu Gupta and Om Prakash Verma. 2022. Monitoring and surveillance of urban road traffic using low altitude drone images: a deep learning approach. *Multimedia Tools and Applications* 81, 14 (2022), 19683–19703.

[11] Javier Hernandez-Ortega, Javier Galbally, Julian Fierrez, Rudolf Haraksim, and Laurent Beslay. 2019. Faceqnet: Quality assessment for face recognition based on deep learning. In *2019 International Conference on Biometrics (ICB)*. IEEE, 1–8.

[12] Glenn Jocher, Alex Stoken, Jirka Borovec, STAN Christopher, and Liu Changyu Laughing. 2021. ultralytics/yolov5: v4. 0-nn. SiLU () activations, Weights & Biases logging, PyTorch Hub integration. *Zenodo* (2021).

[13] Dimosthenis Karatzas, Faisal Shafait, Seiichi Uchida, Masakazu Iwamura, Lluis Gomez i Bigorda, Sergi Robles Mestre, Joan Mas, David Fernandez Mota, Jon Almazan Almazan, and Lluis Pere De Las Heras. 2013. ICDAR 2013 robust reading competition. In *2013 12th International Conference on Document Analysis and Recognition*. 1484–1493. https://doi.org/10.1109/ICDAR.2013.221

[14] Margrit Kasper-Eulaers, Nico Hahn, Stian Berger, Tom Sebulonsen, Øystein Myrland, and Per Egil Kummervold. 2021. Detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5. *Algorithms* 14, 4 (2021), 114. https://doi.org/10.3390/a14040114

[15] Karlo Koščevic and Marko Subašic. 2018. Automatic visual reading of meters using deep learning. In *Croatian Computer Vision Workshop*. 1–6. https://doi.org/10.20532/ccvw.2018.0002

[16] Rayson Laroca, Alessandra B Araujo, Luiz A Zanlorensi, Eduardo C De Almeida, and David Menotti. 2021. Towards image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach. *IEEE Access* 9 (2021), 67569–67584. https://doi.org/10.1109/ACCESS.2021.3077415

[17] Rayson Laroca, Victor Barroso, Matheus A Diniz, Gabriel R Gonçalves, William R Schwartz, and David Menotti. 2019. Convolutional neural networks for automatic meter reading. *Journal of Electronic Imaging* 28, 1 (2019), 013023. https://doi.org/10.1117/1.jei.28.1.013023

[18] Chunshan Li, Yukun Su, Rui Yuan, Dianhui Chu, and Jinhui Zhu. 2019. Lightweight spliced convolution network-based automatic water meter reading in smart city. *IEEE Access* 7 (2019), 174359–174367. https://doi.org/10.1109/ACCESS.2019.2956556

[19] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*. 2980–2988. https://doi.org/10.1109/TPAMI.2018.2858826

[20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

[21] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. 2016. Ssd: Single shot multibox detector.

In *European conference on computer vision*. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

[22] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28 (2015). https://doi.org/10.1109/TPAMI.2016.2577031

[23] Changeui Son, Seokmok Park, J Lee, and Joonki Paik. 2019. Deep learning–based number detection and recognition for gas meter reading. *IEIE Trans. on Smart Processing & Computing* 8, 5 (2019), 367–372. https://doi.org/10.5573/IEIESPC.2019.8.5.367

[24] Chun-Ming Tsai, Tawei David Shou, Shad-Chi Chen, and Jun-Wei Hsieh. 2019. Use SSD to detect the digital region in electricity meter. In *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*. 1–7. https://doi.org/10.1109/ICMLC48188.2019.8949195

[25] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. 2020. CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 390–391. https://doi.org/10.1109/CVPRW50498.2020.00203

[26] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. 2019. PANet: Few-Shot Image Semantic Segmentation With Prototype Alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/ICCV.2019.00929

[27] Yan Chiew Wong, Li Jia Choi, Ranjit Singh Sarban Singh, Haoyu Zhang, and AR Syafeeza. 2019. Deep learning-based racing bib number detection and recognition. *Jordanian Journal of Computers and Information Technology (JJCIT)* 5, 3 (2019), 181–194. https://doi.org/10.5455/jjcit.71-1562747728

[28] Pavinder Yadav, Nidhi Gupta, and Pawan Kumar Sharma. 2022. A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods. *Expert Systems with Applications* (2022), 118698. https://doi.org/10.1016/j.eswa.2022.118698

[29] Pavinder Yadav, Nidhi Gupta, and Pawan Kumar Sharma. 2024. Robust weapon detection in dark environments using Yolov7-DarkVision. *Digital Signal Processing* 145 (2024), 104342. https://doi.org/10.1016/j.dsp.2023.104342

[30] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2018. Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine* 13, 3 (2018), 55–75.