Tech Science Press

Image-Based Automatic Energy Meter Reading Using Deep Learning

Muhammad Imran[1,*], Hafeez Anwar[2], Muhammad Tufail[1], Abdullah Khan[1], Murad Khan[3] and Dzati Athiar Ramli[4]

[1]Institute of Computer Sciences and Information Technology, Faculty of Management and Computer Sciences, The University of Agriculture, Peshawar, Pakistan
[2]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Attock Campus, Attock City, Pakistan
[3]Kuwait College of Science and Technology, Kuwait City, Kuwait
[4]School of Electrical & Electronic Engineering, USM Engineering Campus, Universiti Sains Malaysia, Penang, 14300, Malaysia
*Corresponding Author: Muhammad Imran. Email: mimran@aup.edu.pk
Received: 12 March 2022; Accepted: 02 June 2022

Abstract: We propose to perform an image-based framework for electrical energy meter reading. Our aim is to extract the image region that depicts the digits and then recognize them to record the consumed units. Combining the readings of serial numbers and energy meter units, an automatic billing system using the Internet of Things and a graphical user interface is deployable in a real-time setup. However, such region extraction and character recognition become challenging due to image variations caused by several factors such as partial occlusion due to dust on the meter display, orientation and scale variations caused by camera positioning, and non-uniform illumination caused by shades. To this end, our work evaluates and compares the state-of-the art deep learning algorithm You Only Look Once (YOLO) along with traditional handcrafted features for text extraction and recognition. Our image dataset contains 10,000 images of electrical energy meters and is further expanded by data augmentation such as in-plane rotation and scaling to make the deep learning algorithms robust to these image variations. For training and evaluation, the image dataset is annotated to produce the ground truth of all the images. Consequently, YOLO achieves superior performance over the traditional handcrafted features with an average recognition rate of 98% for all the digits. It proves to be robust against the mentioned image variations compared with the traditional handcrafted features. Our proposed method can be highly instrumental in reducing the time and effort involved in the current meter reading, where workers visit door to door, take images of meters and manually extract readings from these images.

Keywords: Convolutional neural network; object localization; machine learning

1 Introduction

Image-based text detection, extraction and recognition is applied to reduce the time and cost of several applications such as the recognition of number plates [1,2], medical images [3] and road signs [4]. In this regard, the proposed system is based on these techniques by applying them to the novel problem of automatic image-based electrical energy meter reading in Pakistan.

Currently, in Pakistan, meter readers visit door to door and collect the images of electricity meters via smart phones attached to a selfie stick, as shown in Fig. 1. As a normal practice, these meters are placed high on walls or on electricity distribution poles, thus making the whole process of taking images a highly time-consuming task. However, the manual entry of consumed units from these images into a computer is an even more complex, time consuming and laborious task. Our aim in this work is to reduce the effort involved in the second step, namely, the manual entry of the consumed units. We propose a framework that employs an image-based method to recognize the digits on the meter's display to record the meter reading automatically. This is implemented using , state-of-the art deep learning-based object localization frameworks such as You Only Look Once (YOLO) [5] and Faster Region-based Convolutional Neural Networks (Faster RCNN) are utilized. In the current work, we use cropped images to train and test the deep learning model for digits recognition. In real time, the display of the meter is cropped and given to the framework which then applied the trained model to recognize the digits. Consequently, our proposed system is instrumental in reducing the enormous time taken by manual entry of the billing system. Additionally, the system can recognize serial numbers from meters to create an automatic billing system that reads the current units and the user's unique serial number. The framework can be deployed in real-time using the Internet of Things and supported by means of a graphical user interface (GUI).
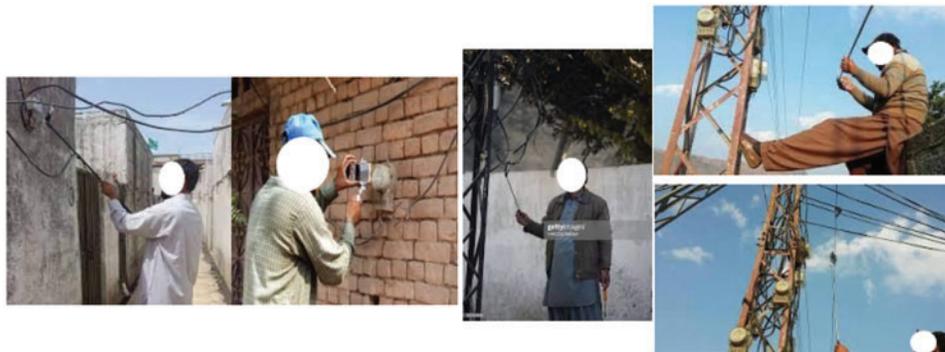

Figure 1: Meter readers taking meter images in Pakistan

Nonetheless, text localization and recognition specifically in a natural environment is very challenging [6–9] due to the inconsistency in imaging conditions [10] and objects' appearance [11]. Electrical meters in Pakistan are placed outdoors so that their images have the same set of variations, as explained in the following and shown in Fig. 2.

- **Image conditions**

a) Shadows due to other objects or reflectance from the glassy surface of the meters cause non-uniform illuminations.

b) Most meters are hung outdoors such as on walls or electric poles so that the dust or mud on their surface cause partial occlusions of the digits.

c) Scale and orientation changes are caused by the position of camera with respect to the meter, which induces variations in the images of the digits as well.
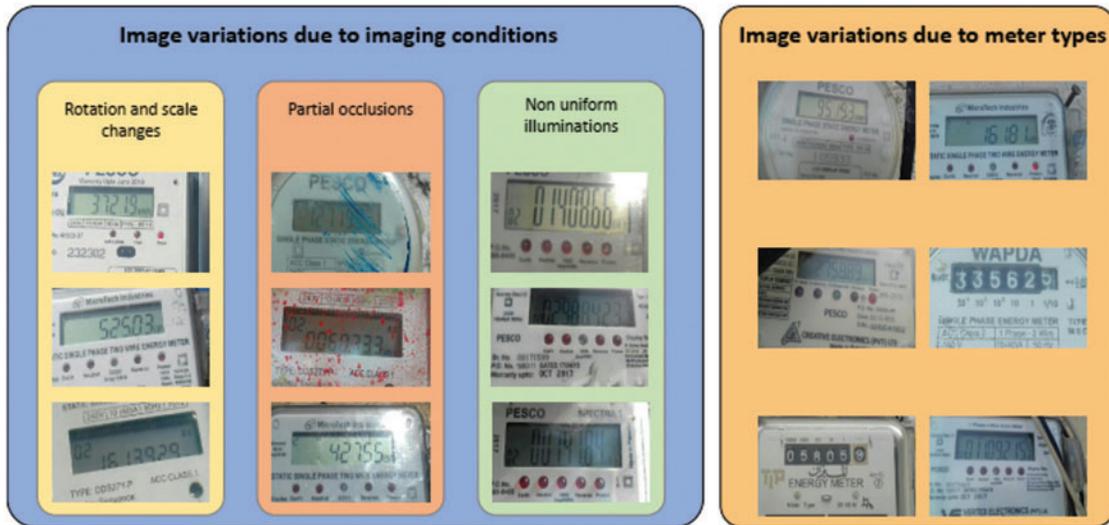

Figure 2: Image variation found in the images of meters

● **Meter types**

The variations in meter types cause the following dissimilarities.

a) The digital display, where the consumed units are shown, varies for different types due to their position and scale.
b) The text fonts are different for each type of meter.
c) The shape of the meter itself induces variations. For instance, some meters are circular, whereas others are rectangular.

Therefore, keeping in view all the said challenges, the main contribution of this paper are as follows.

a) A novel dataset of household electrical energy meter images used in Pakistan is gathered. It contains 10,000 images of five distinct types of meters. All the image variations discussed previously are found in these images.
b) These images are manually annotated to generate their respective ground truths. Annotations are done both for meter displays and then for the digits.
c) The performance of YOLO and handcrafted features is extensively evaluated.

The rest of the paper is organized as follows. Section 2 summarizes the recent method proposed in the literature. Section 3 explains a brief overview of the prepared dataset. Section 4 presents the details of the proposed methodology. Section 5 discusses the achieved results. Finally, Section 6 concludes the paper and states the future directions of our current work.

2 Related Work

Image-based meter reading has obtained less attention from the research community as recently proposed attempts on these applications are few. The two main methods used for this purpose are

YOLO and Faster RCNN. For instance, [12] use FAST-YOLO object detector-based system capable of automatic meter reading. Their approach consists of two steps. The counter is detected in the first step, and the digits are detected in the second step. Their dataset consists of 2,000 pre-annotated images. With the added data augmentation in the dataset, they achieved an accuracy of 97.30% for counter detection and 99.56% for digit detection, outperforming the previously set baselines. Faster R-CNN is used by [13] to solve the problem of image-based meter reading in residential areas. The dataset consists of images of gas and electric meters with 312 and 463 images, respectively. Methods of augmentation are used to extend the dataset to train a robust model. The trained models achieve 100% accuracy in counter detection in images and 99.86% accuracy in serial number detection. Image processing-based method is proposed by [14] for reading the displays of electric meters. Their dataset consists of 21 images of electric meters with the reading in Arabic characters. The accuracy achieved for digit recognition is 82.54% for a single digit of 64×64 pixels and 86.84% for an average of five digits of 64×64 pixels. Another method proposed by [15] uses an image processing approach in meter reading to support the validation of manual reading. The text of the area of interest is extracted using Multi-Net System, which consists of multiple supervised neural models for object detection. This method of neural models locates the object using segmentation. The overall accuracy achieved by the system is 87% on a test dataset of 100 meter counter images. Automatic meter reading using the deep learning approach is performed on a publicly available data set [16] . Their approach is to recognize multidial analog meters along with regular meters. The deep learning architecture used here include Faster R-CNN and YOLO. The F1-score is 100% on dial detection with both architecture, and the accuracy is 93.6% for recognition of dials and 75.25% for meters using Faster R-CNN.

Region-Convolutional Neural Networks (RCNN) is used by [17] to tackle the problem of automatic meter reading. The methods used in this paper include a Mask-RCNN for counter detection, digit segmentation and digit recognition using the a dataset consisting of 2,000 fully annotated images and achieves correct recognition rates of 100%. An image-based meter reading for water meters is proposed in [18] where the robustness to changes in orientation and spatial layout is achieved by enriching the layers of a neural network with necessary information. Image-based meter reading is performed in unconstrained environment [19] via a combination of YOLO-v4 and the so-called (AngReg) regression approach where the meter recognition is 98.90%. A combination of Fast-YOLOv4, corner detection and counter classification network is used by [20] for meter display detection and digits recognition to achieve 99% recognition rate on a dataset of 12,500 meter images.

The summary of papers proposed so far in the literature about image-based meter reading are summarized in Tab. 1, showing the name of the technique used by the paper, the accuracy achieved and the number of images in their datasets. In addition, various applications of YOLO framework are outlined in Tab. 2. Our work is different from them in the following manner.

- The YOLO algorithm is used for the largest dataset of Pakistani meters that contains 10,000 images.
- Our dataset is far more challenging than those datasets because the images in our dataset face a larger set of image variations.

Table 1: Literature about image-based meter reading using deep learning frameworks

Reference no.	Paper title	Algorithm	Accuracy	Dataset
[12]	Convolutional neural networks for automatic meter reading	Fast YOLO Object Detector	83.20% 92%	2000 images

(Continued)

Table 1: Continued

Reference no.	Paper title	Algorithm	Accuracy	Dataset
[13]	Automatic visual reading of meters using deep learning	Faster R CNN	95%	7915 images
[14]	Automatic electricity meter reading based on image processing	OpenCV Library	85.71%	21 images
[15]	Automatic gas meter reading based on image processing	Multi-net	87%	100 images
[16]	Deep learning for image-based automatic dial meter reading: Dataset and baselines	Faster R-CNN YOLO	75.25% 93.6%	N/A
[18]	Image-based automatic water meter reading under challenging environments	YOLOv3	98%	N/A
[19]	Image-based automatic dial meter reading in unconstrained scenarios	YOLOv4	98.90%	5000 images
[20]	Towards Image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach	Fast-YOLOv4	99%	12500 images

Table 2: Various applications of YOLO framework

Reference No.	Paper title	Usage
[21]	A YOLO-based table detection method	Modified YOLO to detect tables in text images
[22]	Cascaded segmentation-detection networks for word-level text spotting	Text detection and recognition in the wild
[23]	Signature and logo detection using deep CNN for document image retrieval	Signature detection and recognition
[24]	UAV-YOLO: Small Object Detection on Unmanned Aerial Vehicle Perspective	Object detection in images from UAV camera
[25]	Face mask wearing detection based on improved YOLO-v4	Detection of masks worn by people in images and videos

3 Dataset

Our dataset consists of the greatest number of images of electrical energy meters of Pakistan. These images are taken by meter reading personnel and then manually uploaded into the billing system. The dataset has 10,000 images of five different types of meters. Some exemplar images of these meters are shown in Fig. 3. The ground truth of these images is generated by annotating them manually. For each of the digits shown in the display section of the units consumed, a bounding box is manually created.

The bounding box for a given digit is drawn in such a manner that it is as close as possible to the digit. This approach ensures the closeness of the ground truth and the predicted bounding box for that digit, hence reducing the margin of errors. Great care is also taken about unambiguous cases during image labelling. An example is a very blurred image where the digits not visible to normal human eye are not labelled. This situation may cause the algorithm to learn incorrect information that can lead to an increase in false detection rate. An example of annotated meter image is shown in Fig. 4.
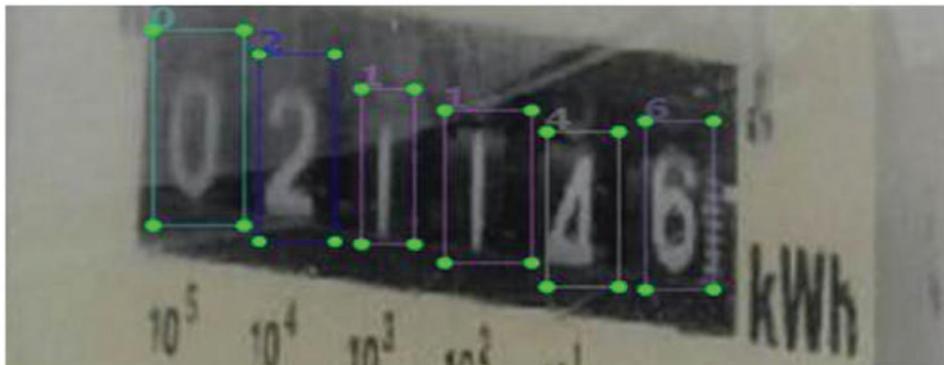

Figure 3: Various types of meters

Figure 4: Image annotation example

4 Methodology

The proposed research methodology is divided into a series of steps to accomplish the proposed research. The steps include i) data collection, ii) data pre-processing iii) and the proposed model. The subsections below provide a brief overview of the research steps.

4.1 Data Collection

Data gathering is a fundamental step in any research area, especially, in artificial neural network [26]. For this paper, 10,000 images from WAPDA office in Pakistan are collected. Our data set contains several types of meters, as shown in Fig. 3, and several types of variations found in those images, as shown in Fig. 2. Our results indicate that the collected pictures present a concrete view of the meter types and variations found for energy meters existing in Pakistan.

4.2 Data Pre-Processing

Data pre-processing is one of the key steps for data scientists [27]. It is used to remove noise and unnecessary information from data, for example images to convert them for efficient processing by neural networks. Many pre-processing steps in literature are applied according to the requirement. In this paper, cropping, resizing, and saturation are applied according to our requirements. Fig. 5a shows the original images before any pre-processing step is applied.
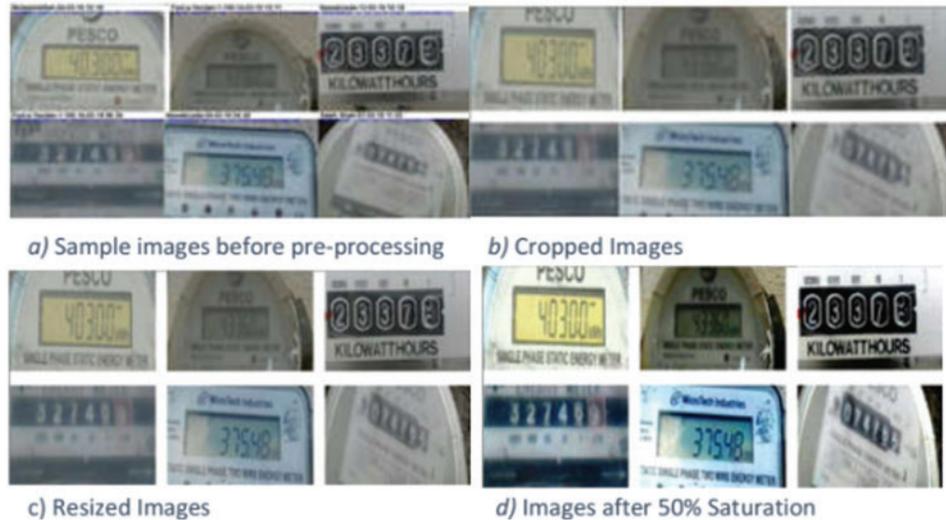
Figure 5: Sample images presenting the results of pre-processing steps

4.2.1 Cropping

Cropping is the removal of unwanted area from images or videos. It is the first step to make raw data ready for processing. In this paper, cropping is used to remove the outer region of energy meter images to separate the subject (meter readings) from unwanted background information. Cropping is generally considered a high-level reasoning for aesthetic enhancement. Fig. 5b shows that the cropped image presents a view of data that can be processed efficiently because unnecessary outer regions are removed from the images.

4.2.2 Resizing

Neural network receives images of the same size for training purposes. Therefore, images must be converted into a fixed size before training. Various image resolutions such as 28×28 pixels, 32×32 pixels and 128×128 pixels or more have been used in the literature. No standard is used in choosing the image size because it mostly depends on the nature of the problem and the implemented architecture. The grade of the details in images is important when resizing. Selecting a smaller size requires less time for training and testing but at the cost of losing important details in images. In this paper, 256×256 pixels is set because at this resolution, the grade of the details is maintained, and it follows the guideline of the underlying architecture, that is YOLOv3. Fig. 5c presents a sample of images after resizing which clearly shows that resizing the images does not affect the quality and important details are maintained.

4.2.3 Saturation

Saturation is a technique that defines the intensity of colour in an image. A lower saturation of a colour is represented by grayscale, whereas a higher saturation is vivid and intense in representation. In this paper, saturation is used to bring darker or bright effects to an image. The saturation of 0% represents a complete dark effect, whereas 100% represents pure white. During experimentation, 50% saturation is selected because it best represents the quality of digits in our dataset. Fig. 5d shows the resultant images that are much sharper than the original images after applying 50% saturation. Fig. 6 summarizes the main blocks of the proposed method.
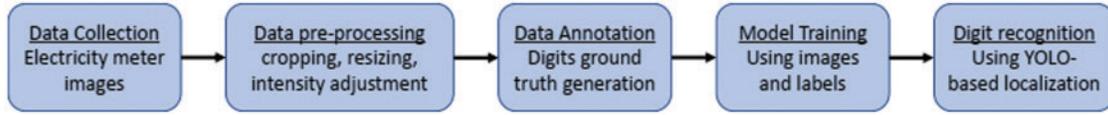
Figure 6: Major steps of the proposed method

4.3 Proposed Model

This work proposes evaluating the traditional handcrafted features-based method and the CNN-based object localization framework named YOLO. Both methods are explained as follows.

4.3.1 Handcrafted Features

This work proposes using the methodology proposed for license plate characters extraction and recognition [1] where the following set of features are calculated to represent the digits in feature vectors.

- **Zoning**: The image patch representing the digit is partitioned into zones, for instance 3×3. The sum of the number of white pixels in each zone is determined.
- **Extent**: This feature is the ratio of white pixels in an image patch to the total number of pixels.
- **Perimeter**: For a connected component, its perimeter is the set of all interior boundary pixels.
- **Orientation**: This feature provides the direction of the major axis of the ellipse around a digit measured with respect to the *x axis*.
- **Sum of rows and columns**: This feature is the sum of the pixel intensities of particular rows and columns of the image patch.
- **Euler Number**: This feature provides the measurement of holes in the digit image.
- **Eccentricity**: This feature measures the circularity of a digit.

With these features, the digits are represented in vectorized forms. Consequently, a support vector machine is used to perform digit recognition.

4.3.2 YOLO Object Detection Algorithm

The input image for the YOLO framework is divided into a grid of $N \times N$ regions. Each of these regions is then answerable for detecting object by predicting a confidence score related to the bounding boxes (B). As shown in Eq. (1), confidence score is the product of object's probability and the intersection over union (IoU) of ground truth and prediction. Consequently, the bounding box with no object has a confidence score of zero; otherwise, it is equal to the IoU of the ground truth and the predicted bounding box.

$$ConfidenceSc = P\,r(Object) \times IoU(truth;\ pred) \tag{1}$$

The YOLOv3 framework is shown in Fig. 7. Each predicted bounding box is parametrized by five values: x, y, w, h and confidence score. The centre point of a predicted bounding box is given by (x, y) coordinates, and its width and height are given by w and h, respectively. For each bounding box 'B', Pr (Class/Object) is denoted by confidence score 'C'. The YOLO architecture is shown in Fig. 8, where the number of layers is 53. Residual blocks from ResNet architecture are added, which increase model accuracy. The input image size to this architecture is $256 \times 256 \times 3$, where the image width and height is 256, and the number of colour channels is 3.
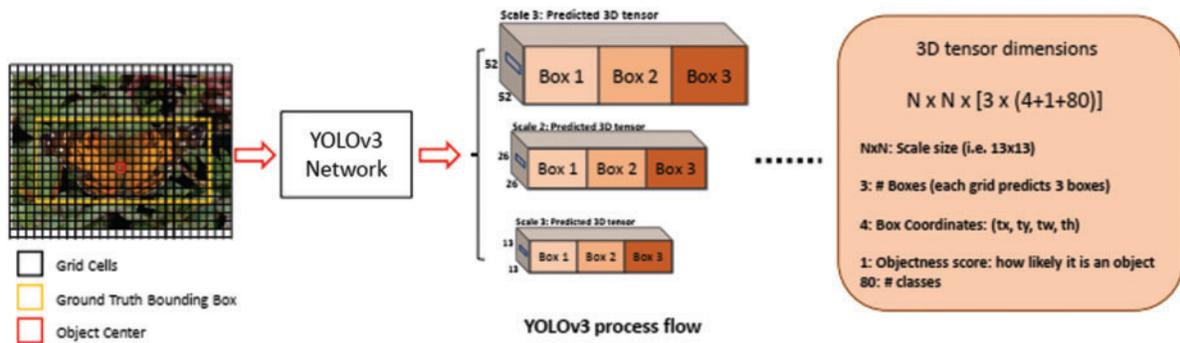
Figure 7: YOLOv3 framework

Layer	Kernel	Padding	Str	Input Size	Output Size
Conv	3x3	same	1	(256,256,32)	(256,256,32)
Conv	3x3	same	2	(256,256,32)	(256,256,64)
Batch Normalization	---	---	---	(128,128,64)	(128,128,64)
Res Block 1x64	3x3	---	---	(128,128,64)	(128,128,64)
Conv	3x3	same	2	(64,64,128)	(64,64,128)
Res Block 2x128	3x3	---	---	(64,64,128)	(64,64,128)
Batch Normalization	---	---	---	(64,64,128)	(64,64,128)
Conv	3x3	same	2	(32,32,256)	(32,32,256)
Batch Normalization	---	---	---	(32,32,256)	(32,32,256)
Res Block 8x 256	3x3	---	---	(32,32,256)	(32,32,256)
Batch Normalization	---	---	---	(32,32,256)	(32,32,256)
Conv	3x3	same	2	(16,16,512)	(16,16,512)
Batch Normalization	---	---	---	(16,16,512)	(16,16,512)
Res Block 8x512	3x3	---	---	(16,16,512)	(16,16,512)
Batch Normalization	---	---	---	(16,16,512)	(16,16,512)
Conv	3x3	same	2	(8,8,1024)	(8,8,1024)
Batch Normalization	---	---	---	(8,8,1024)	(8,8,1024)
Res Block 4x1024	3x3	---	---	(8,8,1024)	(8,8,1024)
Batch Normalization	---	---	---	(8,8,1024)	(8,8,1024)
Conv block	---	---	---	---	---
Batch Normalization	---	---	---	(8,8,1024)	(8,8,1024)

Figure 8: Model architecture

5 Experiments and Results

This section outlines the experimental protocols used for the YOLOv3 framework and the results of the performed experiments.

5.1 Experimental Protocols

The meter image dataset is split into two disjoint sets called the training and test sets. The training set consists of 70% of the images, whereas the test set contains the remaining 30% images. Both the handcrafted features and YOLO framework are trained on the images in the training set along with the respective ground truths, whereas the test set is used to evaluate the performance of the trained model. The system used for experiments has an Intel Core i5 processor of 2.5 GHz with a RAM of 4 GB.

The operating system used is MS Windows 8.1. The backbone CNN architecture used by YOLO is known as DarkNet-53 for digit recognition. As its name suggests, it has 53 convolution layers, and the model has three scale pyramids that are used for various types of object detection. That pyramid is used inside the residual block, which means that it is used for the detection and recognition of all types of images: large, medium, and small. The key parameters of the whole YOLO setup are shown in Tab. 3.

Table 3: Key parameters used for YOLOv3 framework

CNN Architecture	Darknet53
Activation functions	Sigmoid, Leaky ReLu, SoftMax
Loss	Definite cross entropy
No. of epochs	40
Total images	10000
No. training images	7000 images
No. of test images	3000 images
Optimization algorithm	Adam optimizer
Performance metric P1	Classification accuracy
Performance metric P2	Validation loss

5.2 Results

The training and validation losses for 40 epochs are shown in Fig. 9a. Validation loss is jointly caused by classification loss, localization loss and confidence loss. For each cell, classification loss is caused by the detection of incorrect object. The incorrect detection of the location and size of a bounding box contributes to localization loss. The abjectness of a bounding box is measured by confidence loss. Training loss is also observed in the training of the YOLO framework over the training dataset. Training loss is reduced with the increasing number of epochs, but this trend is not observed in validation loss. However, the average validation loss over the complete 40 epochs is 6. The indicative measures of the proposed method for digit recognition are given in Fig. 9b.

(a)   (b)
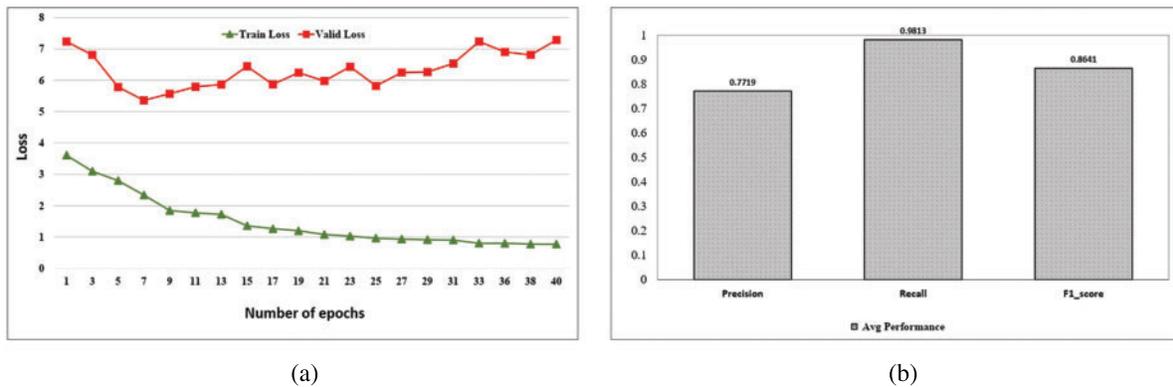
Figure 9: a) Training and validation losses of YOLO with respect to number of epochs b) Various measures indicating the effectiveness of the proposed method on the task of image-based digit recognition

The precision of our method is less than its recall, which means that our proposed method is good at detecting 98% of the actual digits but also detects some of the non-digits as digits, that is false positives. However, in our case, a higher precision is required compared with recall, as we would want the model to detect exactly the same digits that show the energy units consumed. Nonetheless, digit recognition is [14] of utmost importance in meter reading.

The digit recognition rates achieved by YOLO and handcrafted features are shown in Figs. 10 and 11, respectively. YOLO performs better where the recognition is rate is above 98% for most of the digits except '3' and '5'. This finding shows the effectiveness of YOLO for the task of digit recognition. Finally, the detection results of the YOLO framework on some of the meter images are shown in Fig. 12. The incorrect detection of '3' as '9' is shown in the second image of the first row, whereas the digits on the rest of the images are recognized correctly.
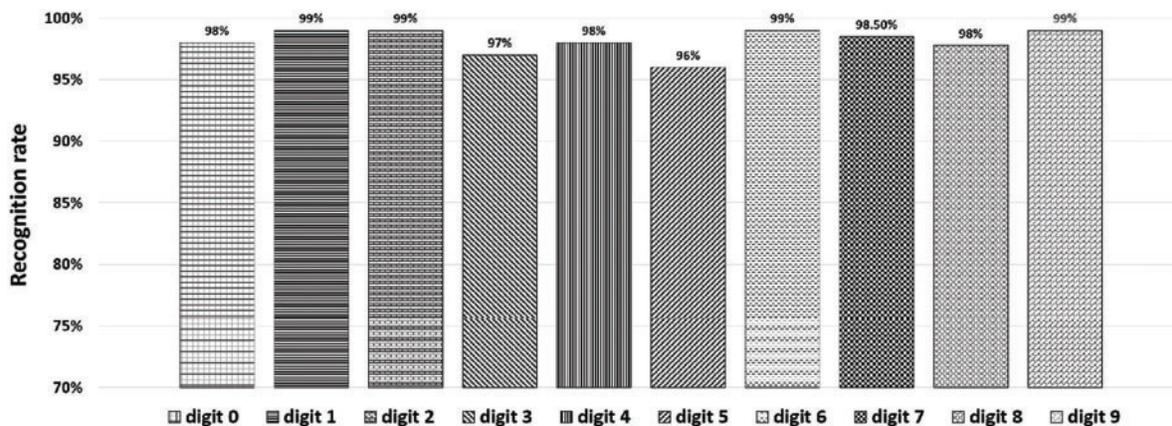

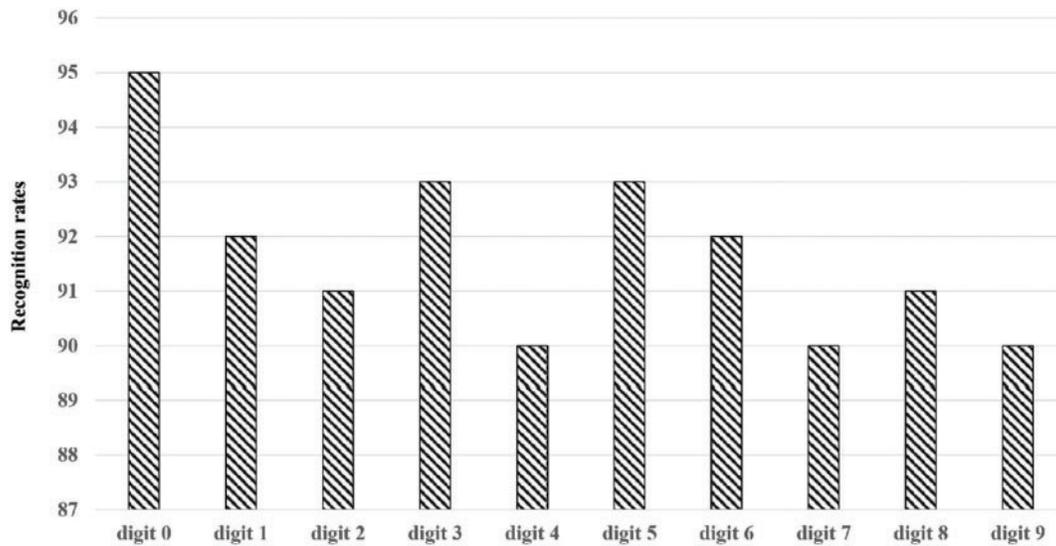**Figure 10:** Recognition rates achieved by YOLO for each digit

Figure. 11: Recognition rates achieved by handcraft feature for each digit
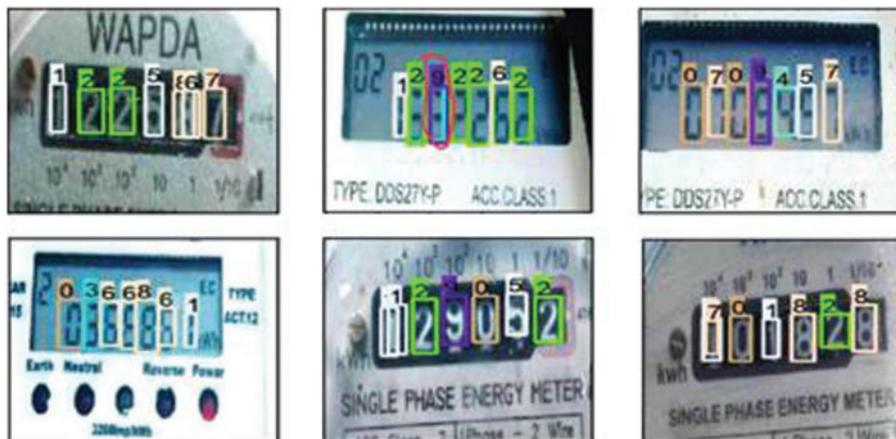

Figure 12: Digit recognition on meter images

6 Conclusion and Future Directions

An automatic meter reading method from images is presented. This method performs text localization, extraction, and recognition by using the YOLOv3 framework. The dataset consists of 10,000 images of electrical energy meters used in Pakistan. These images suffer from variations such as changes in scale and orientation, partial occlusions and non-uniform illumination. In addition, the image variations are caused by the types of meters that differ from one another due to text fonts and placements. With the proposed method, 99% overall digit recognition rate, 77% precision and 98% recall are achieved.

In the future, we plan to extend the dataset by including more types of meters, recognizing the serial numbers from the meters to develop an automatic billing system that records the unit reading and user-specific serial number. The trained models can be deployed into a real-time scenario using a

GUI. The results of the paper can be improved by training models using the upcoming state-of-the-art deep learning architecture and comparing the results to determine which is best suited over the meters dataset by using the same methodology and adding steps such as augmenting and shuffling data to train a robust, more accurate model.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] F. Ullah, H. Anwar, I. Shahzadi, A. U. Rehman, S. Mehmood *et al.,* "Barrier access control using sensors platform and vehicle license plate characters recognition," *Sensors*, vol. 19, pp. 3015, 2019.

[2] B. Jan, H. Farman, M. Khan, M. Imran, I. U. Islam *et al.,* "Deep learning in big data analytics: A comparative study," *Computers and Electrical Engineering*, vol. 75, pp. 275–287, 2019.

[3] W. Xue, Q. Li and Q. Xue, "Text detection and recognition for images of medical laboratory reports with a deep learning approach," *IEEE Access*, vol. 8, pp. 407–416, 2019.

[4] R. Saluja, A. Maheshwari, G. Ramakrishnan, P. Chaudhuri and M. Carman, "OCR on-the-go: Robust end-to-end systems for reading license plates & street signs," in *2019 Int. Conf. on Document Analysis and Recognition (ICDAR)*, Sydney, Australia, pp. 154–159, 2019.

[5] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779–788, 2016.

[6] C. Helwe and S. Elbassuoni, "Arabic named entity recognition via deep co-learning," *Artificial Intelligence Review*, vol. 52, pp. 197–215, 2019.

[7] X. Yuan, C. Li and X. Li, "Deepdefense: Identifying DDoS attack via deep learning," in *2017 IEEE Int. Conf. on Smart Computing (SMARTCOMP)*, Hong Kong, China, pp. 1–8, 2017.

[8] D. Chen, J. M. Odobez and H. Bourlard, "Text detection and recognition in images and video frames," *Pattern Recognition*, vol. 37, pp. 595–608, 2004.

[9] Z. Cheng, F. Bai, Y. Xu, G. Zheng, S. Pu *et al.,* "Focusing attention: Towards accurate text recognition in natural images," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 5076–5084, 2017.

[10] A. Khan, R. Shah, M. Imran, A. Khan, J. I. Bangash *et al.,* "An alternative approach to neural network training based on hybrid bio meta-heuristic algorithm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 3821–3830, 2019.

[11] K. Tang, A. Joulin, L. J. Li and L. F. Fei, "Co-localization in real-world images," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1464–1471, 2014.

[12] R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz *et al.,* "Convolutional neural networks for automatic meter reading," *Journal of Electronic Imaging*, vol. 28, pp. 013–023, 2019.

[13] K. Koščevic and M. Subašic, "Automatic visual reading of meters using deep learning," in *Croatian Computer Vision Workshop*, Zagreb, Croatia, pp. 1–6, 2018.

[14] L. A. Elrefaei, A. Bajaber, S. Natheir, N. AbuSanab, M. Bazi *et al.,* "Automatic electricity meter reading based on image processing," in *2015 IEEE Jordan Conf. on Applied Electrical Engineering and Computing Technologies (AEECT)*, Dead Sea, Jordan, pp. 1–5, 2015.

[15] M. Vanetti, I. Gallo and A. Nodari, "Gas meter reading from real world images using a multi-net system," *Pattern Recognition Letters*, vol. 34, pp. 519–526, 2013.

[16] G. Salomon, R. Laroca and D. Menotti, "Deep learning for image-based automatic dial meter reading: Dataset and baselines," arXiv preprint arXiv:2005.03106, 2020.

[17] A. Azeem, W. Riaz, A. Siddique and U. A. K. Saifullah, "A robust automatic meter reading system based on mask-RCNN," in *2020 IEEE Int. Conf. on Advances in Electrical Engineering and Computer Applications (AEECA)*, Dalian, LiaoNing, China, pp. 209–213, 2020.

[18] Q. Hong, Y. Ding, J. Lin, M. Wang, Q. Wei *et al.,* "Image-based automatic watermeter reading under challenging environments," *Sensors*, vol. 21, no. 2, pp. 434, 2021.

[19] G. Salomon, L. Rayson and M. David, "Image-based automatic dial meter reading in unconstrained scenarios," arXiv, pp. 1–10, 2022.

[20] R. Laroca, A. Alessandra B, Z. Luiz A, D. A. Eduardo C and M. David, "Towards image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach," *IEEE Access*, vol. 9, pp. 67569–67584, 2021.

[21] Y. Huang, Q. Yan, Y. Li, Y. Chen, X. Wang *et al.,* "A YOLO-based table detection method," in *2019 Int. Conf. on Document Analysis and Recognition (ICDAR)*, Sydney, Australia, pp. 813–818, 2019.

[22] S. Qin and R. Manduchi, "Cascaded segmentation-detection networks for word-level text spotting," in *2017 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, pp. 1275–1282, 2017.

[23] N. Sharma, R. Mandal, R. Sharma, U. Pal and M. Blumenstein, "Signature and logo detection using deep CNN for document image retrieval," in *2018 16th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR)*, Niagara Falls, NY, USA, pp. 416–422, 2018.

[24] M. Liu, X. Wang, A. Zhou, X. Fu, Y. Ma *et al.,* "UAV-YOLO: Small object detection on unmanned aerial vehicle perspective," *Sensors*, vol. 20, pp. 2238, 2020.

[25] J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved YOLO-v4," *Sensors*, vol. 21, pp. 3263, 2021.

[26] X. W. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.

[27] X. Zheng, M. Wang and J. Ordieres-Meré, "Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0," *Sensors*, vol. 18, pp. 2146, 2018.