

Automated Meter Reading Detection Using Inception with Single Shot Multi-Box Detector

Arif Iqbal*, Abdul Basit, Imran Ali, Junaid Babar and Ihsan Ullah

Department of Computer Science, University of Balochistan, Quetta, Pakistan

*Corresponding Author: Arif Iqbal. Email: arifiqbalmcs@gmail.com

Received: 09 September 2020; Accepted: 14 December 2020

Abstract: Automated meter reading has recently been adopted by utility service providers for improving the reading and billing process. Images captured during meter reading are incorporated in consumer bills to prevent reporting false reading and ensure transparency. The availability of images captured during the meter reading process presents the potential of completely automating the meter reading process. This paper proposes a convolutional network-based multi-box model for the automatic meter reading. The proposed research leverages the inception model with a single shot detector to achieve high accuracy and efficiency compared to the existing state-of-the-art machine learning methods. We tested the multi-box detector with Mobile-Net and Faster Region-based Convolutional Neural Networks (R-CNN). The results depict that the proposed method not only outperforms the two baseline methods, but also requires less iterations (epochs) to train and rapidly improve precision with 96% accuracy. The dataset used for this research has been collected, preprocessed, and made publicly available to encourage further research and to serve as a baseline for the comparison.

Keywords: Inception model; neural network; object detection; single shot multibox detector

1 Introduction

Sui Southern Gas Company (SSGC) is the leading natural gas supply company in Pakistan. SSGC also owns the only gas meter manufacturing plant in the country. The company bills its consumers on a monthly basis based on the reading of their meters. The staff (also known as meter readers) record the gas consumption by typing the current reading of the meter in a special handheld device and capturing an image of the current meter reading. The process of using these images is a recent advancement in verifying gas bill accuracy, which helps prevent the reporting of false readings. While this process is improving, the information captured during the current process is not being used to its full potential; the images could be used to automatically detect the meter reading. This research attempts to explore the automatic detection of a reading using the captured image. Using automatic detection from meter reading images has many challenges, including the uncontrolled outdoor environment in which the images are captured and the cluster of alphanumeric information surrounding the meter reading, see [Fig. 1](#).



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

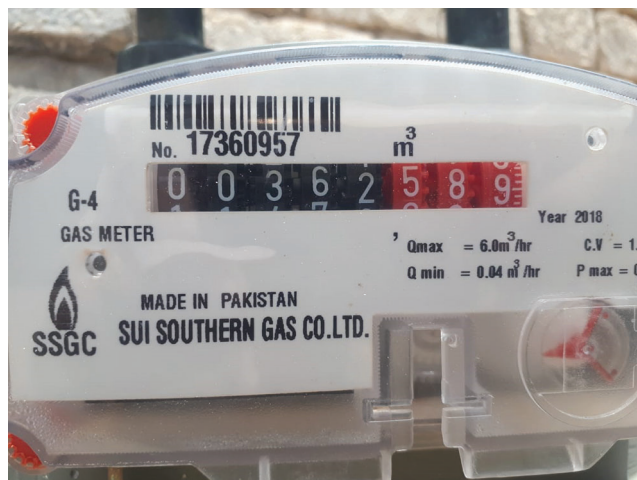


Figure 1: Utility billing meter with analog rotary digit counter. Multiple texts with the key reading region shown with red and black background

These aspects distinguish this text detection application from similar models, and require augmentation of existing techniques proposed in the literature. We propose a model based on deep learning multi-box detector to detect meter readings and extract them in real-time.

2 Related Work

The existing literature discusses various approaches and techniques to detect the meter reading and perform segmentation of the image. In this section, we discuss some of the recent related work on detecting meter readings based on the visual methods.

Elrefaei et al. [1] used image preprocessing techniques to detect the digits counter of electricity and gas meters. The method is comprised of three steps: morphological operations to reduce noise, vertical and horizontal scanning technique for image segmentation and feature extraction with template matching for recognition of digits.

Zheng et al. [2] proposed an algorithm for finding the location of the license plate. The methodology consists of image enhancement, Sobel operator for edge detection, and segmentation of the image to extract the region of the license plate. The proposed algorithm kept track of dark and white pixels horizontally and vertically to receive the license plate region and used a rectangular window larger than the size of the license plate to find the exact location.

Shinde et al. [3] claimed improvements using the Support Vector Machine (SVM) algorithm. In their proposed method, they converted the captured image to grayscale to simplify hue and saturation values. Next, they removed the noise from the image with the help of various morphological operations. A region-based algorithm is used for the extraction of a region of interest (ROI) by scanning high density white and black pixels present around the concerned region. Finally, they used a vertical edge detection algorithm [4] (VEDA) for image segmentation. Vertical edge is a technique that scans the image vertically and marks all the white pixels to find the complete vertical line. The first vertically scanned white pixel is used to create a box for extracting the digit around the scanned area.

Alharaki et al. [5] proposed a neural network-based solution for the recognition of a road tax sticker in Malaysia. They used image processing techniques remove noise from the image and utilized back propagation and the learning vector quantization (LVQ) algorithm for character recognition. The LVQ is a

hybrid network based on supervised learning for classification; the basic steps of the algorithm are initialization, competition, and learning. After removing noise from the image, they scanned it vertically and horizontally to find the location of the road tax sticker. All the characters found on the sticker were classified based on their height and compared. Only the characters with the largest height were selected to segment the image. Finally, they used both back propagation and the LVQ algorithm for character recognition.

Comparably, Vanetti et al. [6] conducted experiments on the gas meter reading detection and recognition by using ensemble-based neural models that were capable of detecting the targeted object. Then, they applied OCR detecting technique for digit recognition.

2.1 Convolutional Neural Network

A convolutional neural network (CNN) [7] is a deep learning algorithm, used to find patterns present in images. The first few layers of the CNN identify the corners and lines before passing these extracted features down to the neural net. As the CNN gets deeper, it is able to begin recognizing more complex features. Basically, the algorithm performs feature learning and image classification.

Cai et al. [8] presented a back propagation (BP) neural network technique for meter reading recognition. The proposed method is comprised of image enhancement techniques, region of interest extraction, and projection-based character segmentation. The segmented character is passed forward to the back propagation neural network as input and then the extracted feature of each character is matched with the assigned weights and bias values. The network back propagates for better recognition of characters.

Jaiswal et al. [9] claimed improvements in the recognition of the digits using image processing techniques and vertical edge detection algorithms for image segmentation. For the digits learning and recognition phase, they used the neural network.

Gomez et al. [10] designed convolutional neural network architecture to perform segmentation free approach for reading text in real-time images. The proposed method comprised of a seven-layer model having a kernel size of 3×3 in all layers and used 2×2 kernels in pooling layers for segmentation and text recognition.

Xie et al. [11] used a convolutional neural network-based MD-YOLO (you only look once) model for the detection of a multidirectional vehicle license plate. To achieve better accuracy over other existing methods, they used angle-based rotation and fast intersection over the union prediction evaluation strategy.

Laroca et al. [12] brought improvements in automatic meter reading by using a convolutional neural network. They designed an approach comprised of two steps. Firstly, they used the Fast-YOLO, a convolutional neural network-based object detector model, for the detection of the digit counter. Secondly, they used three different neural network-based approaches (CR-Net, multitask learning, and recurrent neural network) for digit counter recognition.

Ding et al. [13] used deep learning algorithms on various small data sets of daily items for object detection. CNN was used for feature extraction and classification from images and regional proposal networks (RPN) was used for the final object detection. The model parameters were fine-tuned, enabling the deep learning algorithms to perform faster and with better accuracy.

2.2 Single Shot Object Detector

Liu et al. [14] proposed a deep neural model to detect the various objects. They named the model Single Shot Detector (SSD). The model generates several bounding boxes with different sizes and aspect ratios and discretizes the overlapping bounding boxes into default ground truth boxes. For each of the boxes, the model generates the scores for the object present in the box and adjusts with the box having a similar object shape. It also applies the non-maximum suppression technique to discard overlapping boxes and produce the final

object detection. Furthermore, the network combines multiple extracted features from images having different resolutions to control various size objects. The proposed multi-box model uses VGG-16 [15] as a base architecture to extract features of autonomous objects deeply in real-time environments.

Ning et al. [16] tried to improve the base SSD model by replacing extra layers of the model with the inception block. They claimed their results had been improved by incorporating the inception block as well as by utilizing batch-normalization and non-maximum suppression techniques for more accurate object detection.

Jin et al. [17] used a vanilla single shot detector model for object detection. They replaced a convolution layer with a max pooling layer known as pooling pyramid network (PPN). It is a feature extractor that identifies information from the image and converts the data to a quantifiable format without distorting the image quality. It is an object detection model that is light-weight runs quickly, and performs more accurately when compared to other different single shot detector models.

Tsai [18] used a single shot multi-box detector for digital reading of electricity meters. The pre-trained model worked reasonably well for both collected electricity meter images and the fine-tuned model respectively.

Kim et al. [19] also tried a single shot multi-box detector for the detection and localization of pedestrians and cyclists in the on-road environment. They fine-tuned the model parameters, which increased the overall performance of target object detection and localization compared to the baseline model.

3 Contribution

The major contribution of this research is a single neural network-based model for the detection of the utility meter reading. This research incorporated the inception model with the proposed NN model instead of Mobile-Net and Faster R-CNN. The meter reading detection method based on our proposed method based on our findings is more accurate and faster with 96% accuracy. Fine-tuning the proposed model parameters further improved the training process and the object detection performance.

Additionally, the meter reading dataset has been developed for the first time and has been made publicly available [20]. We tested the dataset with the proposed model and other state of the art methods such as Mobile-Net and Faster R-CNN. The results were significantly improved with the proposed model.

4 Dataset Preparation

The meter reading dataset is not available online, even for research purposes, so formal acquisition of the dataset was not possible. For this research, image datasets of the gas meter readings were collected, preprocessed, prepared, and made available online [20] to facilitate research in the community. The dataset comprises of approximately 1500 images with the resolution of 800×600 in PNG format. We applied morphological operations to remove the noise and resized the images into 300×300 resolutions to improve the training performance.

5 Methodology

After generating the dataset, preparing and preprocessing is the first phase of our proposed model. In this phase, we crop, enhance, and filter the image to suppress the noise. Next, we label each object in the image using the image labeling tool [21] that generates the XML files, which contains the label object information. Later, the object detection model will use the XML files to detect the target object.

The object detection model consists of two phases, the first phase is the inception model and the second is known as the auxiliary network. We use the inception model for feature extraction by applying convolutions of smaller kernel sizes, whereas the auxiliary network performs final object detection, See Fig. 2. Steps are discussed in detail in the next sections.

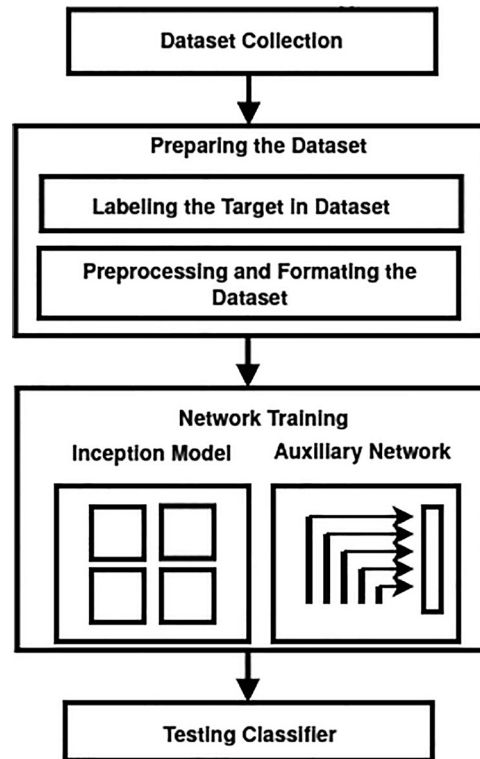


Figure 2: Proposed method block diagram showing various stages of the algorithm. After gathering the dataset, we labeled the target area in the training dataset and applied formatting and preprocessing. Next, we carried the training of the dataset with the two-phase model, inception model, and an auxiliary network. Finally, we test the pre-trained classifier for the object detection

5.1 Data Pre-processing and Formatting

Our first step was to record the target object information that we wanted our model to train on and later detect in the test phase. To do this, we labeled the target object throughout the dataset, using the graphical annotation tool that stores the object information into the XML file. For each image, the XML file contains the position (x; y), width, height, and the class of the object.

Each image with a resolution of 800×600 was resized to 300×300 pixels to increase the training performance of the model. Then, we applied various image morphological operations to remove noise in the images. The model further converts the XML files into the comma-separated values (CSV) file format. Because the model is unable to process the CSV file format as an input, we needed to convert CSV into a TFRecord file format. This binary file format is easier and faster to process and load during the training phase, so the TFRecord format optimized the data feed. Finally, the label `map.pbtxt` file is generated which maps every object class label to an integer value and is used in both training and detection phases.

5.2 Proposed Network Architecture

After preprocessing and formatting the image and target object in the dataset, we fed the image to the training model. The training model took the input image and moved it forward to the first convolutional layer, where we resized it in order to feed it forward to the next convolution layers of different filter sizes for feature extraction of the target object as well as, to perform deeper final object detection.

The inception model consists of 42 deep convolutional layers. The computational cost is higher than Google-LeNet [22] and is more efficient than VGG-Net [23]. The model uses different-size convolution kernels that are added to the input image with a maximum pooling layer that decreases computational costs by reducing the number of parameters.

We treated the outputs of the feature extraction with the pooling layer and concatenated them all to feed it forward to the next convolutional layers. This allowed the model to take advantage of multilevel extraction of the features at global 3×3 and local 1×1 of the target object simultaneously without affecting the network complexity and with improved performance. After concatenation of the features of the target object, the inception model passed these extracted features to the auxiliary network.

The auxiliary network performed object detection by generating more convolutional layers at the end of the first part of the model and decreasing the size of kernels at multiple scales: 10×10 , 5×5 , 3×3 and 1×1 . This helps the model predict objects on multiple scales.

5.3 Fine-tuning Parameters

The single shot multi-box detector benchmark architecture was further fine-tuned using the Inception model to make the model more robust on meter reading detection. The initial learning rate varies in the range of 0:002–0:004. The batch size value of 12 is used for the sliding window.

We maintained the value of the aspect ratios for use with object localization. Since the input image size required 300×300 pixels, the original images were resized from 800×600 into 300×300 . We observed improved accuracy, reported in detail in the results.

5.4 Training Phase

The default boxes are matched with ground truth boxes to train the model. Then, each ground truth box is matched with the best-overlapped default box and those bounding boxes are selected for the final object detection whose intersection over union (IOU) is greater than the 0:5 threshold. The model requires a minimum of 100 and a maximum of 500 images for accurate detection of the target object. This training is faster and less data is required. The overall training loss is defined in terms of confidence and localization loss is expressed in equation no (1).

$$L(x; c; l; g) = 1/N(L_{\text{conf}}(x; c) + \alpha L_{\text{loc}}(x; l; g)) \quad (1)$$

where (L) denotes the overall loss, (x) represents the input image, (c) stands for confidence, (l) is used for predicted box, and (g) for ground truth boxes. Whereas, (N) denotes the number of default matched boxes, the localization loss (Lloc) depends on smoothing the input image and confidence loss (Lconf) based on the softmax classifier. Typically 80% of the dataset is used for training. The training is performed for 50 epochs with 5,000 steps per epoch. Once the training phase is completed, the proposed model generates checkpoint files and configuration files used during the training. Finally, the inference graph is exported for validation purposes, See Fig. 3.

6 Experiments and Results

After training the proposed model with our dataset, we set up a challenging environment to test its robustness. We collected a large number of images containing gas meter reading with different sensors, lighting conditions, resolutions, and viewpoints. We tested both the speed and accuracy of the proposed method and compared it with the baseline methods.

To establish a comparable baseline, we also tested the dataset with two famous neural network-based models: Mobile-Net and Faster R-CNN. Two experimental setups for testing were built. In the first experiment, we created a dataset of randomly collected images. We applied both the proposed and the

baseline methods over the dataset and calculated the performance. In the second experiment, we linked the training phase with the Tensor Board (a web application used for monitoring the precision and recall of a model in real-time) to record the performance of the model.

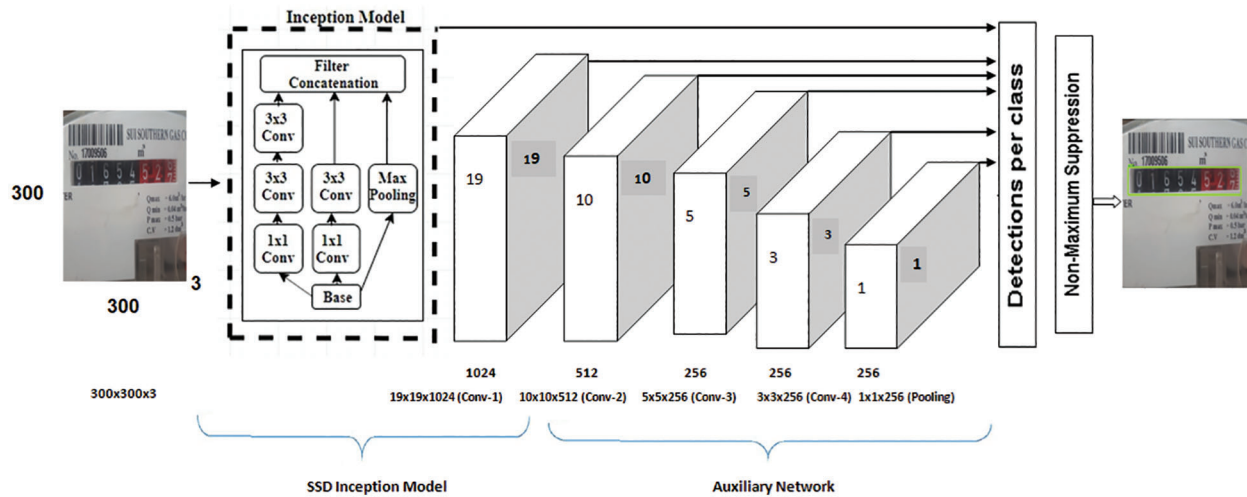


Figure 3: The proposed network model with various layers used to train over the dataset. The proposed network model uses the inception model where we used different size kernels with a max-pooling layer and concatenated all extracted features simultaneously without affecting the network complexity. After concatenation, the extracted features feed-forward to the auxiliary network that generates more convolutional layers and decrease kernel sizes at multiple scales to perform deeper final target object detection

6.1 Experiment I

One hundred (100) images were randomly collected from various sources, the images captured in varying daytime with different camera sensors, viewpoints, and scales.

As already mentioned, we conducted the training of the model on the Google GPU to save the time, but we conducted the tests both on local CPU and Google GPU. For testing the model locally, we downloaded the trained inference graph.pb extension, a binary file that is used for final object detection and label map.pbtxt file that contains the labeling information (x; y), width, height, and class of the target object.

The configuration of the local CPU is Intel Core i3 1.61 GHz processor 8GB RAM and used Python Jupyter notebook on Linux environment whereas, the GPU is Google Colab (Python3 Tesla K80) to examine the robustness of our proposed model in terms of accuracy and speed.

The proposed method processed 12.86 fps to detect the target object on the CPU, whereas the method processed 41.63 fps on the graphical processor (GPU). The baseline methods Mobile-Net processed 7.22 fps on CPU and 25.53 fps on the GPU, the Faster R-CNN processed 0.64 fps on CPU and 8.27 fps on the GPU.

Considering the accuracy, the proposed method detects the target with 96% accuracy, whereas the baseline methods Mobile-Net performs with 92% and the Faster R-CNN with 89% accuracy, outperforming the baseline methods by 4% and 7% respectively, see [Tab. 1](#).

In addition to outperforming the baseline methods in accuracy, the proposed method outperformed both baseline methods with a significant margin both in speed. The proposed method improves about 6 and 13 frames over both the baseline methods when tested on CPU, whereas it shows 16 and 33 frames improvement over GPU. We tested the proposed method over the collected image dataset; the outcome of the method detected the object correctly with significant accuracy, See [Fig. 4](#).

Table 1: Proposed method comparison with baseline methods

Model	Speed		Accuracy (%)
	CPU	GPU	
Inception	0.078 (Fps:12.86)	0.078 (Fps:12.86)	96
	0.024 (Fps:41.63)	0.024 (Fps:41.63)	
Mobile-net	0.139 (Fps:7.22)	0.139 (Fps:7.22)	92
	0.039 (Fps:25.53)	0.039 (Fps:25.53)	
Faster R-CNN	1.568 (Fps:0.64)	1.568 (Fps:0.64)	89
	0.121 (Fps:8.27)	0.121 (Fps:8.27)	

The proposed method outperforms the mobile-net by 4% whereas, the faster R-CNN by 7% inaccuracy. The proposed method also outperformed the baseline methods in terms of speed both on the CPU and GPU.

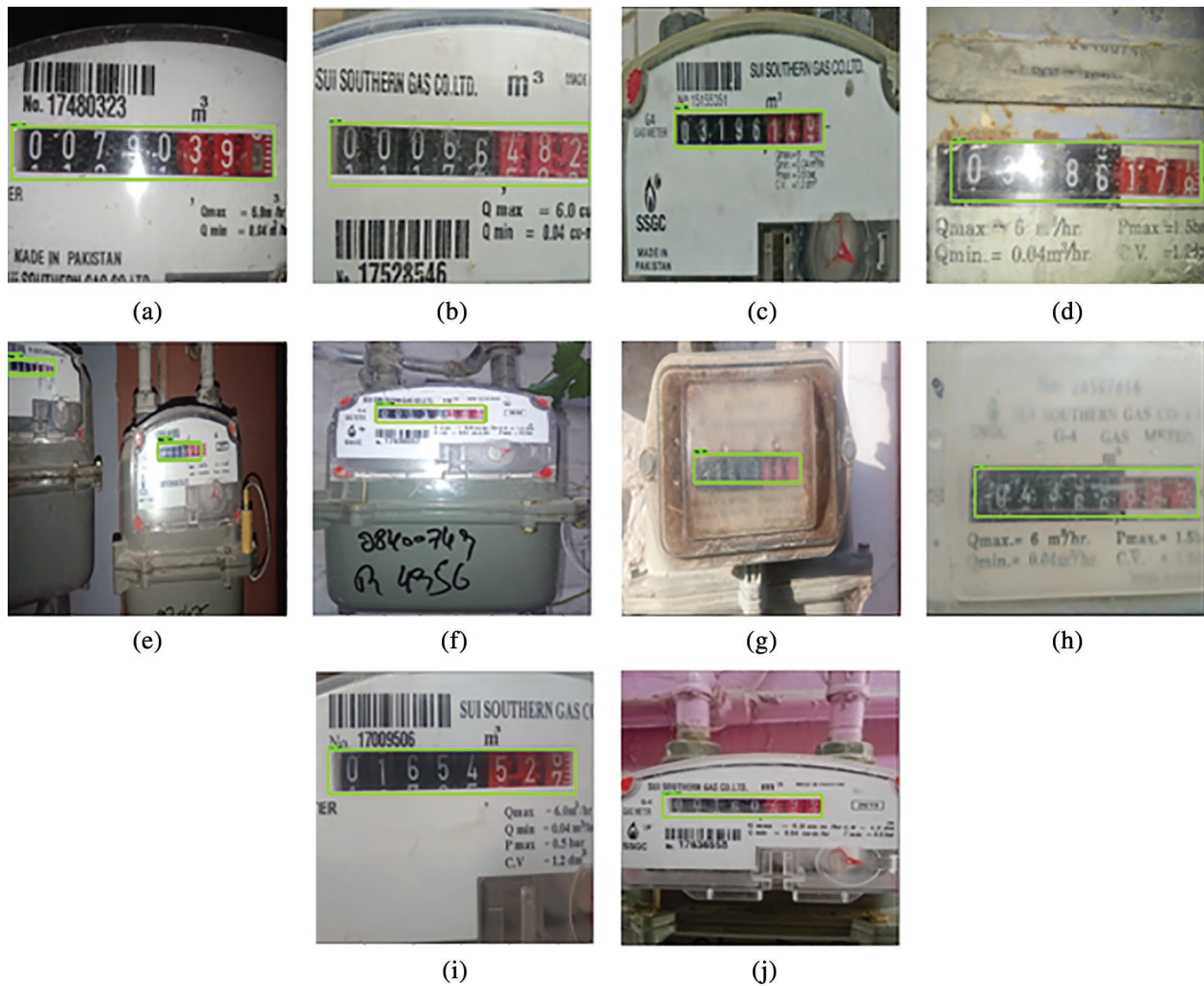


Figure 4: Gas meter reading detection with the proposed method in varying outdoor environments. In (b), (c), and (i), the method performs remarkably well even with the existence of other text including bar-code. (a), (d), (e), and (f) the method performs in a strong illumination effect. In (e) the method detects the multiple detection. In (j) the method performs in a color image

6.2 Experiment II

In experiment II, we carried out additional tests to ensure the robustness of the proposed method. We used Tensor board, the web-based precision evaluation tool, to check the precision of the model during the training phase. The board operates by reading Tensor Flow event files generated during the training phase to visualize the performance. The training parameters were linked to the board which updates precision and recall graphs during the training depending upon the number of training iterations.

Using the Tensor board, we were able to visualize the precision of the proposed and the baseline methods. The proposed method precision improves quickly from 800 epochs and reaches to 96% after 4,500 epochs, the baseline Mobile-Net method improves slowly and reaches to 92% after 15,000 epochs, whereas the second baseline method improves slowly and remains stable at 85% for a longer time and finally reaches to 89% after 20,000 epochs, See Fig. 5.

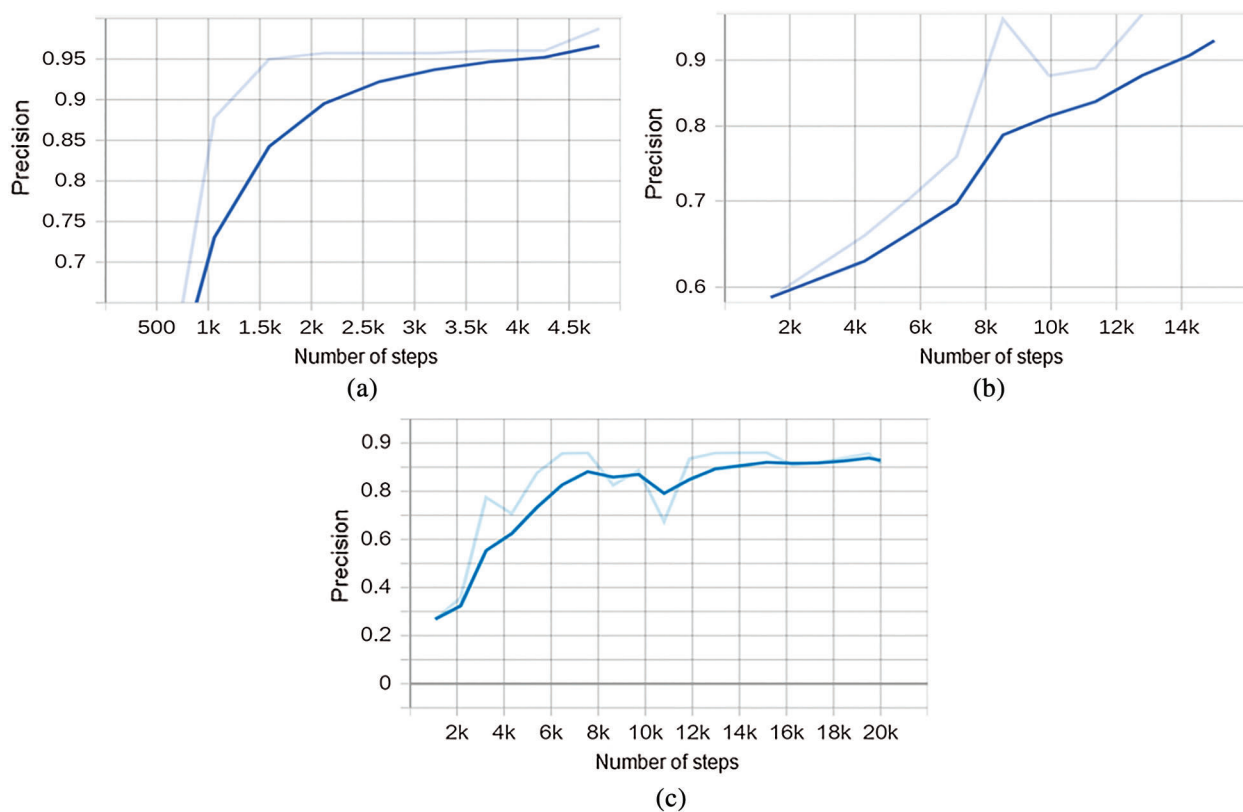


Figure 5: Accuracy graph generated by the Tensor Board. (a) The proposed method (b) Baseline Mobile-Net (c) Baseline Faster R-CNN

The proposed method for the meter reading requires fewer epochs and present improved results and outperforms the baseline methods.

7 Conclusion

We proposed a fast and accurate neural network-based model for automatic meter reading based off of the images that are already being gathered in the course of the gas billing process. We conducted various experiments and the results showed that the proposed method outperforms the baseline methods both in

accuracy and speed. The proposed method additionally trains quickly with fewer epochs than the baseline methods and shows 96% accuracy.

In the future, we plan to expand our work to accurately segment the detected text and apply the character recognition methods.

Funding Statement: The author(s) received no specific funding for this study.

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

References

- [1] L. A. Elrefaei, A. Bajaber, S. Natheir, N. Abu Sanab and M. Bazi, "Automatic electricity meter reading based on image processing," in *Jordan Conf. on Applied Electrical Engineering and Computing Technologies (AEECT)*, IEEE, Amman, pp. 1–5, 2015.
- [2] D. Zheng, Y. Zhao and J. Wang, "An efficient method of license plate location," *International Journal of Pattern Recognition Letters*, vol. 26, no. 15, pp. 2431–2438, 2005.
- [3] M. Shinde and P. Kulkarni, "Reading of energy meter based on image processing technology," *International Journal of Electronics Communication and Computer Engineering*, vol. 5, no. 4, pp. 1–5, 2014.
- [4] A. M. Al-Ghaili, S. Mashohor, A. Ismail and A. R. Ramli, "A new vertical edge detection algorithm and its application," in *Int. Conf. on Computer Engineering Systems*. IEEE, Cairo, pp. 204–209, 2008.
- [5] O. O. Alharaki and A. M. Zeki, "Image recognition technique of road tax sticker in Malaysia," in *Int. Conf. on Advanced Computer Science Applications and Technologies (ACSAT)*. IEEE, Kuala Lumpur, pp. 397–401, 2012.
- [6] M. Vanetti, I. Gallo and A. Nodari, "Gas meter reading from real world images using a multi-net system," *International Journal of Pattern Recognition Letters*, vol. 34, no. 5, pp. 519–526, 2013.
- [7] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, 25, Lake Tahoe, Nevada, USA, F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger, eds., vol. 60, Curran Associates Inc., pp. 1097–1105, 2012.
- [8] Z. Cai, C. Wei and Y. Yuan, "An efficient method for electric meter readings automatic location and recognition," *International Journal of Procedia Engineering*, vol. 23, pp. 565–571, 2011.
- [9] S. Jaiswal, A. Prajapati, N. Shiroadkar, P. Tanawade and C. Gala, "Electricity meter reading based on image processing," *International Journal of Engineering Science and Computing*, vol. 7, no. 4, pp. 11189–11190, 2017.
- [10] L. Gomez, M. Rusinol and D. Karatzas, "Cutting Sayre's knot: Reading scene text without segmentation. Application to utility meters," in *13th IAPR International Workshop on Document Analysis Systems (DAS)*. IEEE, Vienna, pp. 97–102, 2018.
- [11] L. Xie, T. Ahmad, L. Jin, Y. Liu and S. Zhang, "A new CNN based method for multi-directional car license plate detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 507–517, 2018.
- [12] R. Laroca, V. Barroso, M. A. Diniz, G. R. Gonçalves, W. R. Schwartz *et al.*, "Convolutional neural networks for automatic meter reading," *International Journal of Electronic Imaging*, vol. 28, no. 1, pp. 013–023, 2019.
- [13] S. Ding and K. Zhao, "Research on daily objects detection based on deep neural network," *IOP Conf. Series: Materials Science and Engineering*, vol. 322, no. 6, pp. 024–062, IOP Publishing, Nanjing, China, 2018.
- [14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed *et al.*, "Single shot multibox detector," in *European Conf. on Computer Vision*. Cham: Springer, vol. 9905, pp. 21–37, 2016.
- [15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computing Research Repository (CORR)*, arXiv:1409.1556, vol. 9905, 2015.
- [16] C. Ning, H. Zhou, Y. Song and J. Tang, "Inception single shot multibox detector for object detection," in *IEEE Int. Conf. on Multimedia & Expo Workshops (ICMEW)*. IEEE, Hong Kong, pp. 549–554, 2017.
- [17] P. Jin, V. Rathod and X. Zhu, "Pooling pyramid network for object detection," *Computing Research Repository (CORR)*, arXiv:1807.03284, vol. 1, 2018.

- [18] C. M. Tsai, T. D. Shou, S. C. Chen and J. W. Hsieh, "Use SSD to detect the digital region in electricity meter," in *Int. Conf. on Machine Learning and Cybernetics (ICMLC)*. IEEE, Kobe, Japan, pp. 1–7, 2019.
- [19] H. Kim, Y. Lee, B. Yim, E. Park and H. Kim, "On-road object detection using deep neural network," in *IEEE Int. Conf. on Consumer Electronics-Asia (ICCE-Asia)*, IEEE, Seoul, pp. 1–4, 2016.
- [20] A. Iqbal and A. Basit, "Meter reading dataset for machine learning. 2019. [Online]. Available: <https://github.com/arifiqbal2018/meter>.
- [21] Y. S. Yun, J. Jung, S. Eun, S. S. So and J. Heo, "Detection of GUI elements on sketch images using object detector based on deep neural networks," in *Int. Conf. on Green and Human Information Technology*, Springer, Singapore, pp. 86–90, 2018.
- [22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Boston, pp. 1–9, 2015.
- [23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, (CVPR)*, IEEE, Las Vegas, pp. 2818–2826, 2016.