Received: 10 June 2023 | Revised: 4 September 2023 | Accepted: 10 October 2023 | Published online: 24 October 2023

RESEARCH ARTICLE

Artificial Intelligence and Applications
2023, Vol. 00(00) 1–12
DOI: 10.47852/bonviewAIA32021197

An Attention-Based Fusion of ResNet50 and InceptionV3 Model for Water Meter Digit Recognition

Lama Alkhaled[1], **Ayush Roy**[2,*] and **Palaiahnakote Shivakumara**[3]

[1]*Department of Computer Science, Electrical and Space Engineering, Lulea Tekniska Universitet, Sweden*

[2]*Department of Electrical Engineering, Jadavpur University, India*

[3]*Faculty of Computer Science and Information Technology, University of Malaya, Malaysia*

Abstract: Digital water meter digit recognition from images of water meter readings is a challenging research problem. One key reason is that this might be a lack of publicly available datasets to develop such methods. Another reason is the digits suffer from poor quality. In this work, we develop a dataset, called MR-AMR-v1, which comprises 10 different digits (0–9) that are commonly found in electrical and electronic water meter readings. Additionally, we generate a synthetic benchmarking dataset to make the proposed model robust. We propose a weighted probability averaging ensemble-based water meter digit recognition method applied to snapshots of the Fourier transformed convolution block attention module-aided combined ResNet50-InceptionV3 architecture. This benchmarking method achieves an accuracy of 88% on test set images (benchmarking data). Our model also achieves a high accuracy of 97.73% on the MNIST dataset. We benchmark the result on this dataset using the proposed method after performing an exhaustive set of experiments.

Keywords: water meter digit recognition, FCBAM, MR-AMR dataset, computer vision

1. Introduction

The amount of freshwater is reducing day by day due to pollution caused by humans. If water is not used responsibly, the possibility of acute water shortage will be an inevitable reality. Water as a natural resource is an essential demand for not only the daily purposes of humans but also for industries, agriculture, etc. Due to this, there is water wastage. The utilization of emerging technologies like Internet of Things (IoT), artificial intelligence (AI), etc. integrated with traditional methods is used to automate and limit wastage.

It has been repeatedly demonstrated that AI/machine learning (ML), in conjunction with IoT, is the way forward. Practical implementations in prominent fields, such as healthcare [1], environmental assessments [2], the defense industry [3], document classification and verification [4, 5], and scene understanding [6, 7], are just a few examples. We believe that AI/ML and IoT can be effectively utilized in monitoring and managing the world's most valuable natural resource, water. Water wastage is a significant problem that can have catastrophic consequences for humanity. To address this issue, automated water meters are now in use to track usage and reduce wastage. However, some customers have reported receiving water invoices that are significantly higher than their usual monthly payments, sometimes reaching up to $150. Users can utilize data from their meters to identify and address any issues with their excess water consumption. AI/ML, IoT, and computer vision (CV) applications have been used extensively in this field to read water meter data. However, this challenging task requires a dedicated dataset to effectively tackle the problem. In the same way, the automatic system of water meter reading can be reduced to human intervention and labor work. The same work can also be used to recognize handwritten digit (MNIST dataset). Therefore, the proposed system is useful for many real-world applications such as bank cheque verification, authentication, person identification, personality identification, and automation.

To tackle these problems, we create the largest and most complex water meter digit recognition dataset, the MR-AMR dataset, and provide a robust benchmarking technique (augmented images to mimic the real-life challenges) for testing the models in the future. Also, the work focuses on creating novel attention-aided combined convolutional neural network (CNN) architecture with ensemble techniques to provide benchmarking results on the dataset for researchers to beat.

The key contributions of this work are as follows:

- We develop a database, dubbed MR-AMR-v1, which comprises 10 classes (digits from 0 to 9) and is the largest existing and most complex water meter digit recognition dataset.

*Corresponding author: Ayush Roy, Department of Electrical Engineering, Jadavpur University, India. Email: aroy80321@gmail.com

- We generate a synthetic dataset by applying some image processing methods to the original images for robust benchmarking of the models.
- We propose a snapshot ensemble method applied to a Fourier transformed convolution block attention module (FCBAM) attention-aided combined ResNet50-InceptionV3 CNN architecture for classifying the digit images.
- We benchmark the results on the datasets using the proposed method after performing an exhaustive set of experiments.

The remaining part of this article is organized as follows: Section 1 describes some previous methods that performed circuit component recognition. The preparation of the dataset that is made publicly available is described in Section 2. Sections 3 and 4 discuss the model training, benchmarking, and experimentations. Finally, the paper is concluded in Section 5.

2. Related Works

In recent years, researchers have been exploring various approaches to address the challenges in water meter digit recognition. Among the proposed methods, the use of CNNs has gained significant attention. For instance, in Veerai-ah and Chandrashekara [8], the authors propose an automated water meter reading system that applies image processing techniques to recognize the digits on the meter. Similarly, in Wei et al. [9], an automatic water meter reading system using ML techniques is proposed. In Zhang et al. [10], a deep learning-based approach based on a CNN is presented for a water meter reading. Furthermore, in Reddy and Ray [11], different digit recognition techniques, including neural network-based approaches, are compared for an automated water meter reading. Finally, in Rashid and Islam [12], an edge detection and Hough transform-based method are proposed for an automated water meter reading. These methods highlight the potential of using ML and image processing techniques, including CNNs, for accurate and efficient water meter digit recognition. Nonetheless, there is still room for further research and improvement in this area, and future work can focus on exploring new techniques and algorithms to achieve higher accuracy and efficiency in automated water meter reading systems.

Transfer learning has become a widely adopted strategy in the field of CV due to its effectiveness in challenging domains. ResNets and Inception models are well-established and commonly utilized in this specific area. Numerous research papers have suggested applying transfer learning to various domains [13], including water meter digit recognition systems. Hu et al. [14] propose a system using a ResNet-based deep CNN, while Ghanbari et al. [15] propose an automated system using a ResNet-50 deep learning model. Zhang et al. [16] and Liu and Wang [17] propose systems based on ResNet-based residual networks with a region proposal network. Wang et al. [18] propose a system using ResNet-based residual networks and transfer learning to improve the performance of the network. Similarly, Park et al. [19] propose a system based on the Inception-v3 network with data augmentation, while Zheng and Zhang [20] propose a system using the Inception-v4 network. Zhang et al. [21] propose a system using a combination of the Inception-v3 network and a deep CNN. Zhou [22] proposes a system using the Inception-v4 network with data augmentation. In all these systems, a digital camera is used to capture the image of the water meter, and then the respective network is applied to recognize the digits on the meter. These systems have shown promising results, demonstrating the effectiveness of transfer learning and deep neural networks in water meter digit recognition.

Researchers have found success in enhancing the performance of water meter digit recognition systems by combining two modes. Guo [23] proposes a system that utilizes the Inception-ResNet-v2 architecture for deep learning-based water meter reading. The system captures the image of the water meter using a digital camera and then applies the network to recognize the digits on the meter. Similarly, Yan et al. [24] propose a water meter reading system that uses a combined Inception-ResNet-v2 network. Liu et al. [25] propose a system based on a combined ResNet and Inception network, while Zhang et al. [26] propose a system that uses a combined Inception-v4 and ResNet50 network for a water meter reading. Sun et al. [27] propose a system that uses a combined Inception and ResNet network with ensemble learning for water meter reading using a digital camera to capture the image of the water meter and then apply the network to recognize the digits on the meter.

To enhance model performance, researchers have explored various techniques such as combining multiple modes, attention mechanisms, and snapshot ensembles. The convolutional block attention module (CBAM) is a popular attention model that has been used to improve the performance of models. Woo et al. [28] showed the effectiveness of CBAM on digit recognition using the MNIST dataset, outperforming other attention mechanisms, including the squeeze-and-excitation module. Other researchers have proposed using CBAM in conjunction with other techniques. Luo et al. [29] proposed a dual attention network (DAN) for multispectral pedestrian detection that uses both spatial and channel attention mechanisms and applied the FCBAM module to the DAN for digit recognition on the MNIST dataset. Nakanishi and Matsumoto [30] also used CBAM in convolutional layers of their model for digit recognition and showed improved performance on the MNIST and USPS datasets. Zhang et al. [31] proposed an attention-based capsule network (ABCNet) for handwritten digit recognition, which uses both spatial and channel attention mechanisms, and applied CBAM to the ABCNet. Huang et al. [32] introduced the snapshot ensemble technique, which significantly improves deep neural network performance, and applied it to various image classification tasks, including digit recognition using the MNIST dataset. However, CBAM focuses on local features, so to enhance its focus on global features, we have utilized the global frequency domain features using fast Fourier transform (FFT)-based feature as input for the CBAM. In a related paper, DeVries and Taylor [33] proposed a variant of snapshot ensembles called “stochastic weight averaging with restarts”, which involves taking multiple snapshots of the model during training and showed improved performance on various image classification tasks including digit recognition. Finally, He et al. [34] introduced the Residual Network (ResNet) architecture, which has become a popular choice for image classification tasks. The authors applied ResNet to various image classification tasks, including digit recognition using the MNIST dataset, and achieved state-of-the-art results.

In recent years, researchers have made various attempts to create a water meter digit recognition dataset for training CNNs. Li et al. [35] proposed a water meter reading system based on image processing techniques, such as image segmentation, feature extraction, and classification, achieving high accuracy in recognizing water meter digits. Zhang et al. [36] proposed a deep learning-based water meter reading system that can recognize water meter digits in real-time, using a CNN to extract features from water meter images and achieve high recognition accuracy.

Wang et al. [37] compared the performance of several deep learning models, including CNN, recurrent neural network (RNN), and convolutional long short-term memory (LSTM), for water meter digit recognition, with CNN and convolutional LSTM outperforming RNN for this task. Chen et al. [38] proposed a water meter digit recognition system based on histogram-based features and support vector machines, achieving high accuracy in recognizing water meter digits and outperforming other traditional ML algorithms. Sakthimohan et al. (2023) also developed a model for MINIST handwritten digit recognition using CNN. Keerthi [39] proposed CNN-based approach for MINIST handwritten digit recognition. Masthan et al. [41] explored CNN in different way for recognizing MINIST handwritten digits. The main focus of these method is limited to handwritten digit particularly MINIST dataset. Therefore, the methods may not be effective for water meter digit recognition. However, the lack of a dataset with a larger number of images that matches the complexity of real-time noise, distortions, etc., has motivated the creation of the MR-AMR dataset.

3. Proposed Methodology

ResNet50 and InceptionV3 are two widely used CNN architectures in CV tasks. In the proposed approach, the input image of dimension (160 × 160) is fed to both ResNet50 and InceptionV3. The last feature map of dimension $C \times H \times W$ from each network is then taken, where C is the depth and H and W are the width and height of the feature map, respectively. C is 2048 for both ResNet50 and InceptionV3. For ResNet50, the feature map has a size of 5 × 5, while for InceptionV3, it has a size of 3 × 3. FCBAM attention is then applied to these feature maps to enhance their informative regions and suppress the noise, ultimately improving the performance of the model in recognizing water meter digits. Similar to CBAM, FCBAM considers the feature map F as a fast Fourier transformed feature of the input feature map ($F = FFT\{Fin\}$, Fin = output feature map of DenseNet121). Thus, a frequency domain feature helps us in global feature attention. A 1D channel attention module (CAM) and a 2D spatial attention are applied to the Fourier transformed feature map F of DenseNet121 of dimension $C \times H \times W$. CAM essentially provides weights to channels of the feature maps, i.e., enhances particular channels, which contribute more toward boosting the model performance. The 1D channel attention network outputs a feature map Fc of dimensions C × 1 × 1.

$$Fc = s(MLP\{GAP(F)\} + MLP\{GMP(F)\}) \tag{1}$$

where s = sigmoid activation function and + is the addition operation (the two feature maps from global average pooling (GAP) and global max pooling (GMP) are added). The MLP layer comprising two dense layers is shared by GAP layer and GMP layer with a relu activation function.

The first dense layer has output dimensions of C/r (r = reduction ratio) followed by the second layer of output dimension C.

Now, Fc', i.e.,

$$Fc' = Fc \times F \tag{2}$$

is fed the spatial attention module (SAM). SAM is the domain space encapsulation attention mask applied to the feature maps to enhance the features. SAM outputs a feature map F” of dimensions $C \times H \times W$.

$$F'' = f^{7\times7}[MLP(GAP(Fc')); MLP(GMP(Fc'))] \tag{3}$$

where $f^{7\times7}$ is the convolutional layer of kernel size 7 × 7 with dilation of 4 and ; denotes the concatenation of the two feature maps.

$$Fout = F'' \times Fc' \tag{4}$$

Now this $Fout$, which is the output of the FCBAM attention module having dimensions $C \times H \times W$, passes through a GAP layer to reduce its dimensions to C. This C dimensional 1D feature from both the baseline CNNs is added together and fed to the classification layer. The pictorial demonstration of FCBAM is shown in Figure 1(b). Our classification MLP layer has one dense layer of output dimensions 10 with a sigmoid activation function. Snapshot ensemble is used to create 2 snapshots while training the model. We develop the snapshot ensemble in two parts – the first part involves a custom call-back to save the model at the bottom of each learning rate schedule while the second part involves loading the saved models and using them to make an ensemble prediction. Here, we are using cosine annealing learning rate scheduling. We have implemented the cosine annealing schedule as described in Yan et al. [24]. The equation of cosine equation is as follows:

$$w(t) = \frac{wo}{2}\left(cos\left(\frac{t-1, \lceil N/M \rceil}{\lceil N/M \rceil}\right) + 1\right) \tag{5}$$

Here, [] represent the floor function, N is the total number of training epochs, M is the number of cycles, the mod is the modulo operation, w0 is the maximum learning rate, and w(t) is the learning rate at epoch t. The confidence scores from the 2 saved snapshots are averaged giving more weightage to the best snapshot.

The predicted class is the index of the modified confidence score matrix corresponding to the highest value. Let the confidence score array of snapshot 1 and snapshot 2 be CF1 and CF2, respectively, where snapshot 1 is the best model. The weighted average confidence score matrix CFavg is

$$\mathrm{CFavg} = (0.6 \times log(CF1) + 0.4 \times log(\mathrm{CF2})) \tag{6}$$

The predicted class I can be found using CF3:

$$I = arg \max_{x \to c} CFavg(x) \tag{7}$$

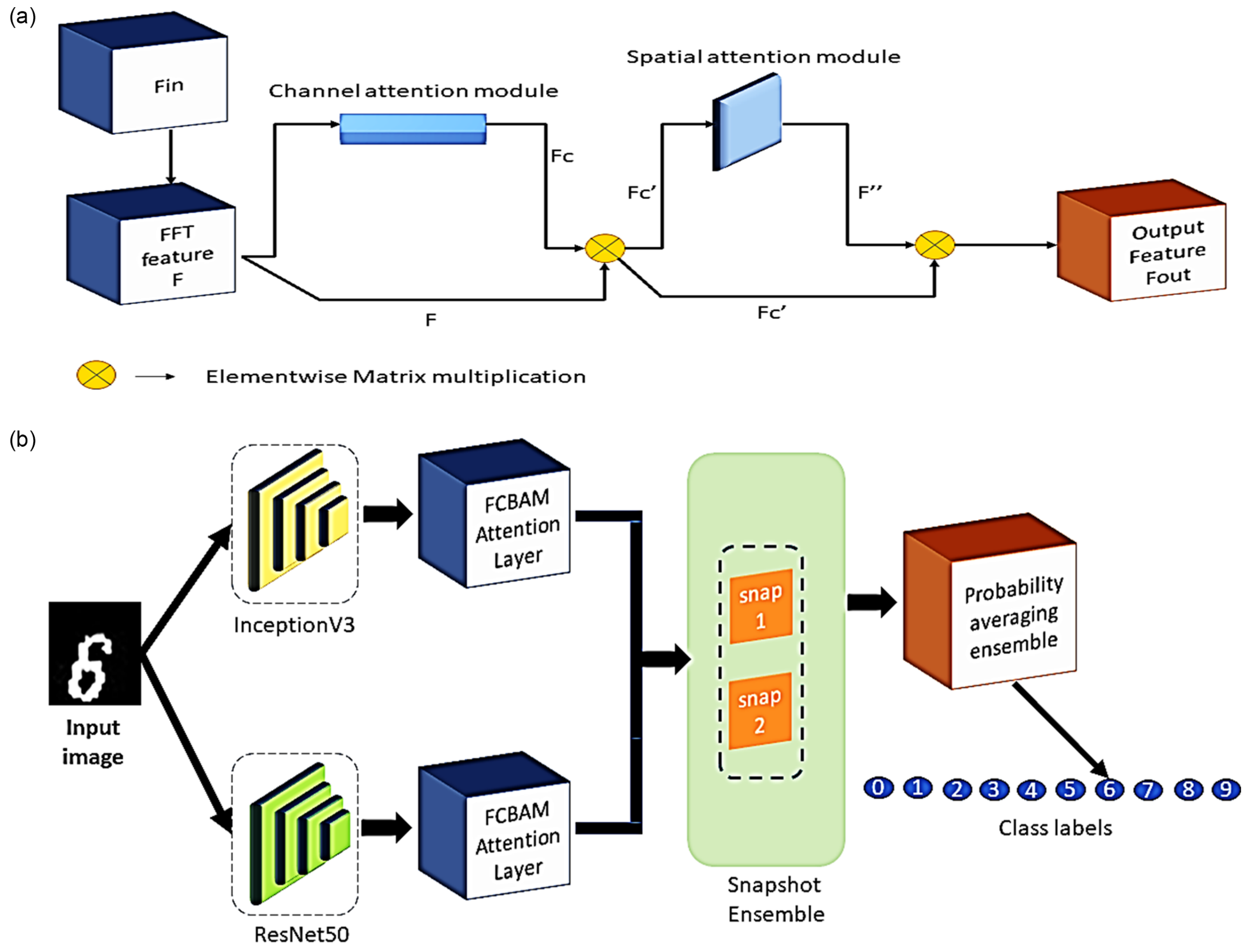
where $c = 1,2,3, \ldots\ldots, C$ ($C = 10$ is the total number of classes). To normalize the confidence scores, the logarithmic operation is applied.

The overall block diagram of the proposed model is shown in Figure 1(a).

3.1. Model training

The models were trained for a total of 50 epochs using a fixed learning rate of 0.001 across all experiments. To implement the snapshot ensemble technique, each cycle was set to 25 epochs, which generated two snapshots for the entire training duration of

Figure 1
(a) The block diagram representation of the proposed model. (b) A pictorial demonstration of CBAM attention module



50 epochs. The training curves for the proposed model can be seen in Figure 2.

4. Experimental Results

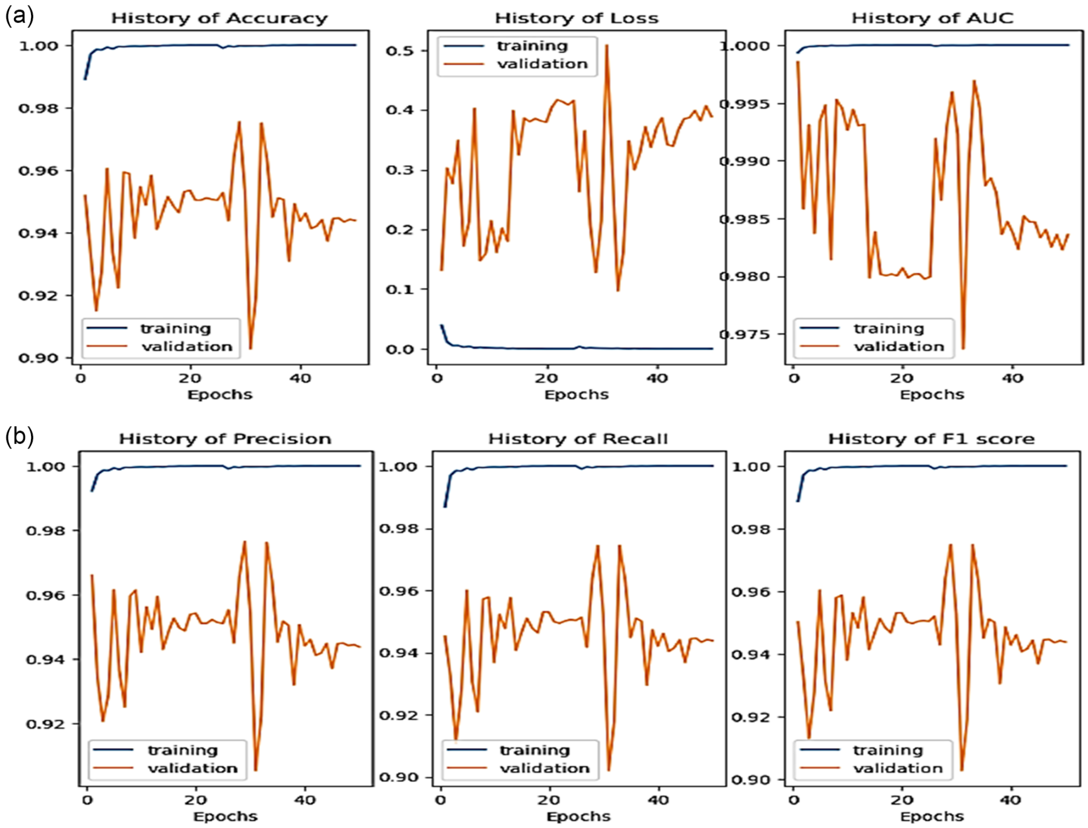
This section presents different augmentation techniques to increase the number of samples and discusses several experiments to evaluate the proposed method. In addition, it also presents ablation study with different baselines and well-known architectures to show that the proposed architecture is superior to other architectures.

4.1. Dataset collection

The MR-AMR dataset [40] is a comprehensive dataset for water meter digit recognition, comprising approximately 140,000 digit meter images collected from various Moroccan meters. The images were captured using different Canon cameras in RGB format with a resolution of 5184 ×3456. To extract the region of interest, the images are cropped by parameterizing the four new coordinates of the required rectangular area, followed by Gaussian smoothing to filter the cropped image and reduce noise. The smoothed image is then converted to grayscale, and each grayscale image is binarized using the adaptive threshold mean binarization function. To segment the individual digits, the binarized image is divided into eight equal divisions, as a maximum of eight numbers is present in a water meter reading. Each digit image is then resized to a dimension of 28×28 pixels.

The dataset is created using a semi-automated annotation format to ensure proper distribution of images per class, and it includes a wide variety of all possible states for a particular class of image, making the dataset complex for digit recognition and similar to real-world cases. The authors have included a complete pipeline for creating the dataset, as shown in Figure 4. In addition, the authors have augmented the images for testing purposes to make benchmarking on the dataset more challenging, as described in Section 2.2. Overall, the MR-AMR dataset is the largest and most complex dataset for water meter digit recognition, and it provides a realistic and challenging environment for developing and evaluating deep learning models for this task.

Figure 2
The training curves for the proposed model



4.2. Data augmentation and evaluation

Ten samples from the original MR-AMR dataset are taken. Four different augmentation techniques are used on these images:

- Clockwise rotation: The original images are rotated by an angle.
- Anticlockwise rotation: The original images are rotated by an angle.
- Brightness diminution: The brightness of the original image is decreased by 40%.
- Binary pixel addition: Converting a few black background pixels into the white foreground/ character pixels in a randomized manner.

A few samples of the augmented images are shown in Figure 5. All these images (10 original + 40 augmented images per class) are saved in the folder "challenge". This folder is used for testing the robustness of the model for low-quality degraded images (these images mimic the problems faced in real-time applications where the images may get rotated or become noisy). Sample images of the augmented images are shown in Figure 6.

To benchmark the results on our dataset as well as on the MNIST dataset, we have used accuracy, precision, recall, and f1 score for the evaluation of the models.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{8}$$

$$Precision = \frac{(TP)}{(TP + FP)} \tag{9}$$

$$Recall = \frac{(TP)}{(TP + FN)} \tag{10}$$

$$F1 - score = 2 \times \frac{(Precision \times Recall)}{(TPrecision + Recall)} \tag{11}$$

where TP = true positive class predictions, TN = true negative class predictions, FP = false positive class predictions, and FN = false negative class predictions. Adam optimizer is used with a learning rate of 0.001 (wo in Equation (5)) for training the models. Cosine annealing is important in optimization algorithms, especially for training deep neural networks, as it helps to improve convergence and prevent overfitting by smoothly varying the learning rate. The learning rate is reduced gradually following a cosine curve, allowing the model to explore different regions of the loss landscape and find better local minima, which can lead to better generalization and more accurate models. The tensorflow-keras library is used for the implementation of the models while the Open-CV, PIL, and os libraries are used to access and augment the images. For the evaluation metrics, we utilized the inbuilt functions of the Sklearn library of Python.

Figure 3
Confusion matrix of the two generated snapshot ensemble models on MR-AMR dataset

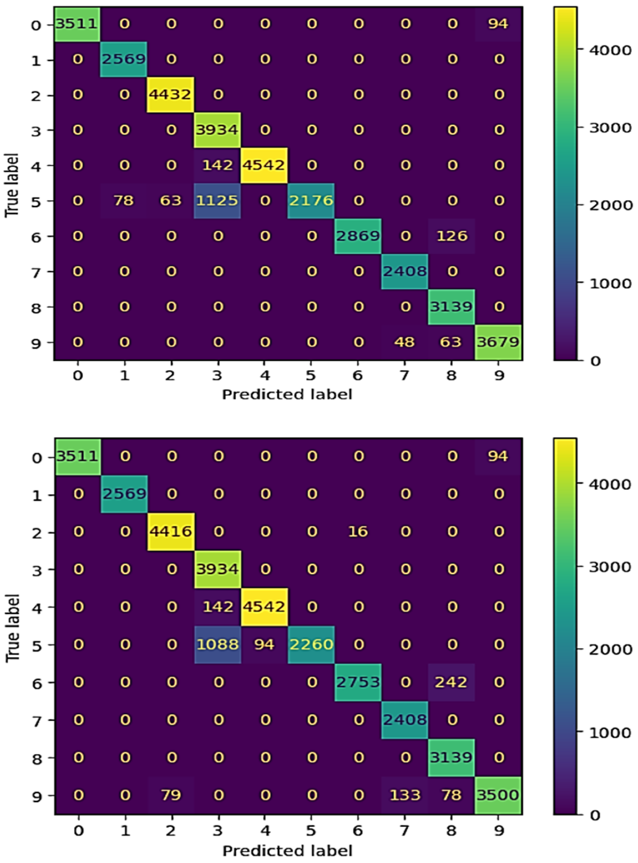


Figure 4
Pipeline for data collection for the MR-AMR dataset

4.3. Ablation study

We have performed the ablation study on the MRAMR dataset and the best-performing model from the ablation study is used for the MR-AMR benchmarking dataset to generate benchmarking results.

ResNet50: ResNet50 achieved an accuracy of approximately 94% on the MR-AMR dataset (see Figure 7), by training 1/3 of

Figure 5
A comparison between the complexity of MNIST and MR-AMR datasets. From the images, it is seen that MR-AMR is a highly challenging dataset and ideal for training models to be installed in water meters as real-world use cases

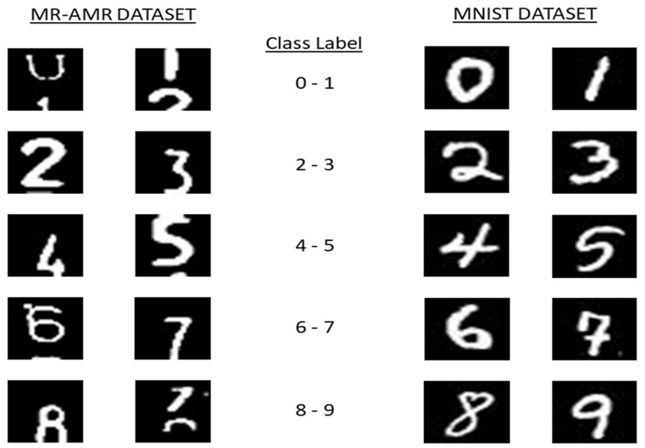
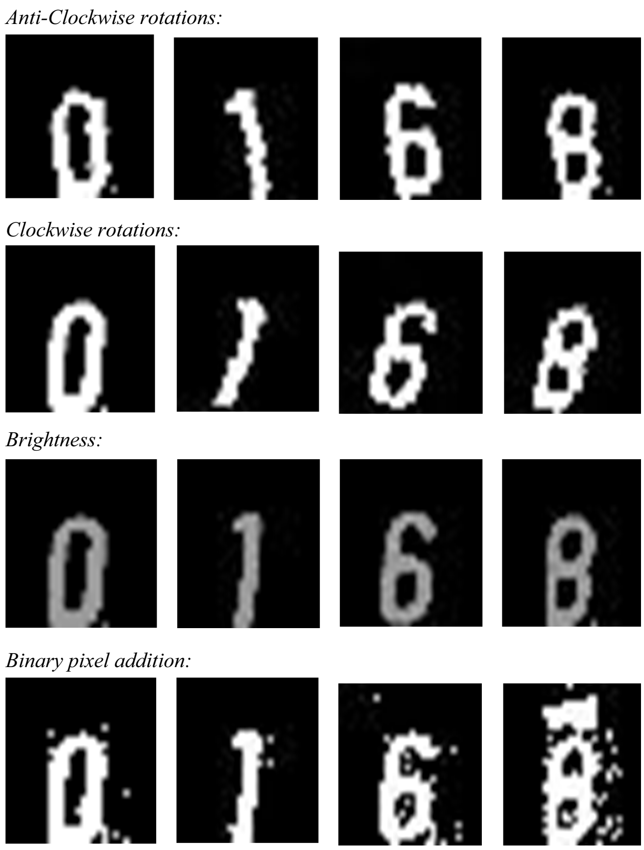


Figure 6
The above images represent the augmented testing dataset for benchmarking purpose. The testing dataset is made challenging by using augmentations that degrade the images to test the robustness of the model



the layers (i.e., the last 50 layers) while freezing the remaining layers with pre-trained ImageNet weights. The training was conducted for 50 epochs, and two snapshots were saved (one at the 25th epoch and the other at the 50th epoch) for the snapshot ensemble.

InceptionV3s: In the experiments conducted, the InceptionV3 model achieved an accuracy of 94.15%, as shown in Figure 7. The model was trained for 50 epochs with 1/3 of the layers (i.e., the last 100 layers) being trainable, while the remaining layers were frozen with pre-trained ImageNet weights. The training process involved saving 2 snapshots, with each snapshot being taken after 25 epochs (for each cycle of the snapshot ensemble).

FCBAM applied on ResNet50 + InceptionV3: In Section 4.3 of our study, we observed that the accuracy improved after combining the ResNet50 and InceptionV3 models. To further enhance the performance of our model, we applied the FCBAM attention mechanism on the last layers of both models before merging their features (as shown in Figure 1(a)). The performance of the improved model is illustrated in Figure 7, where we achieved an accuracy of 95.03%, which is a significant improvement from the accuracy of 94.87% obtained by combining ResNet50 and InceptionV3 without FCBAM. This demonstrates the effectiveness of FCBAM in improving the performance of our water meter digit recognition model.

Weighted probability averaging ensemble: To further improve the accuracy of our model, we utilize the classification ensemble technique called the weighted probability ensemble technique, as shown in Equation (5). The increase in accuracy and decrease in loss on the MR-AMR dataset are illustrated in Figures 7 and 8, respectively. The parameters of the models are listed in Table 1. The confusion matrix of the two best snapshots of the proposed model is shown in Figure 3.

Table 1
The parameters of the models

Model	Total parameters	Trainable parameters	Non-trainable parameters
ResNet50	23,608,202	16,971,018	6,637,184
InceptionV3	21,823,274	13,644,554	8,178,720
ResNet50 +InceptionV3	45,410,986	30,595,082	14,815,904
ResNet50 +InceptionV3 +FCBAM	47,508,336	32,692,432	14,815,904

4.4. Experiments on recognition

Figure 9 shows performance of the proposed method for classification of digits in terms of confusion matrices on the main dataset and benchmark challenging dataset. The strong diagonal dominance in the confusion matrix demonstrates the high accuracy for recognition of our proposed model. Our model shows strong performance on standard datasets such as MNIST

Figure 7
The comparison results of the proposed model on MR-AMR dataset

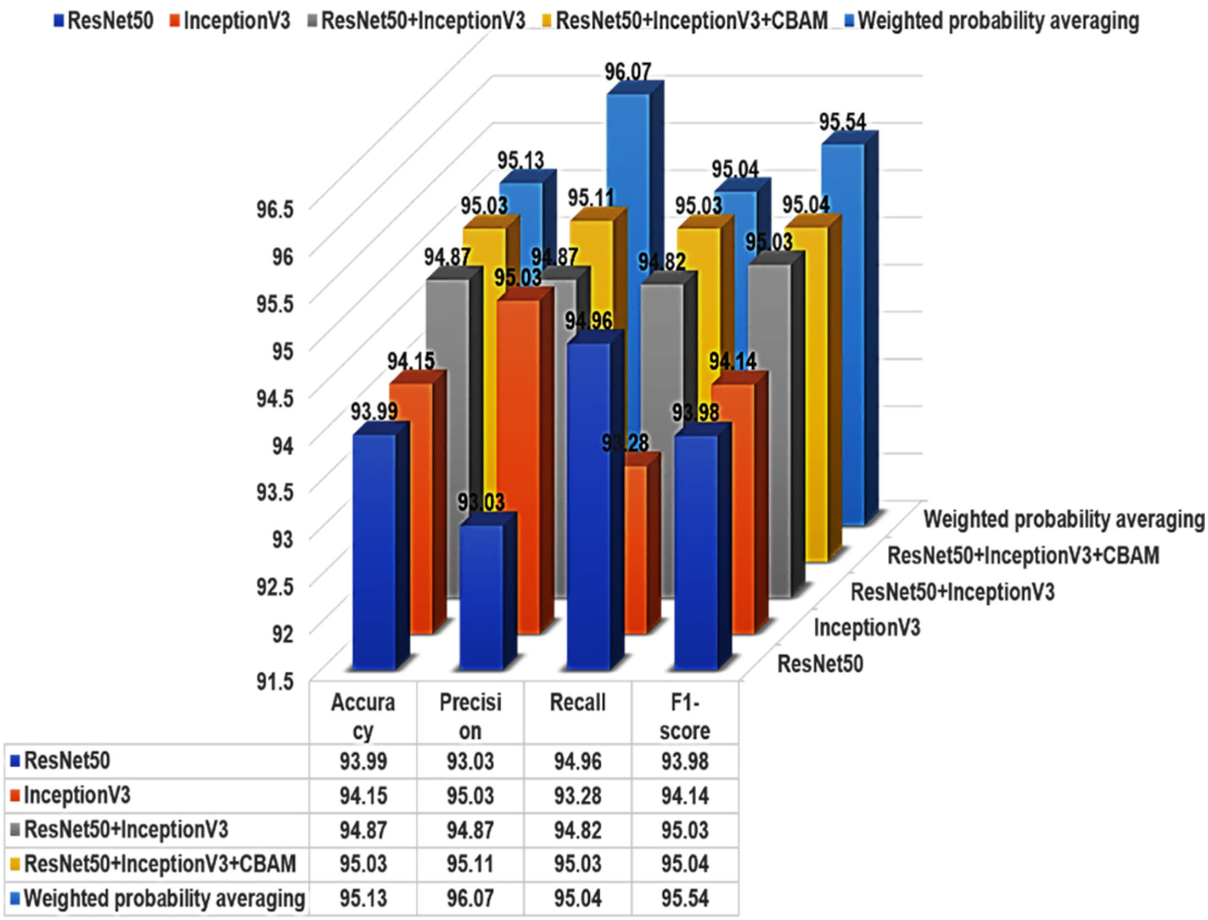


Figure 8
Illustration of the validation loss of the models

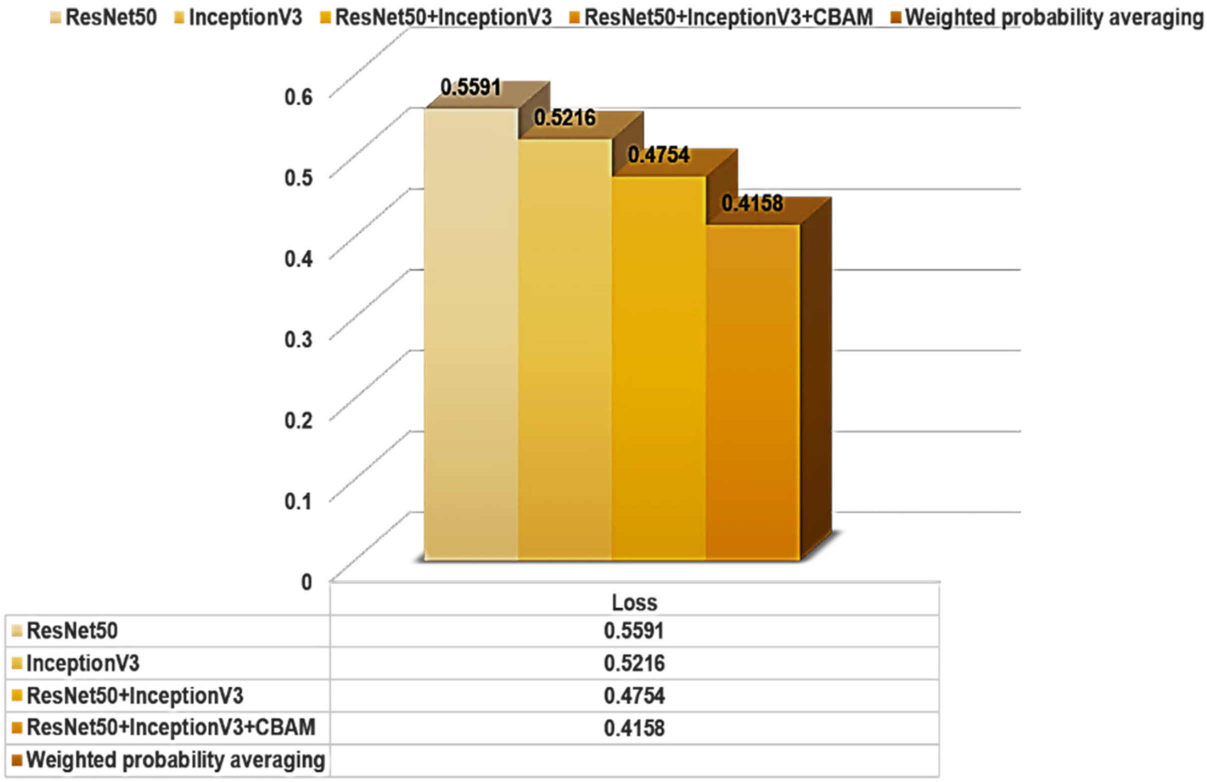
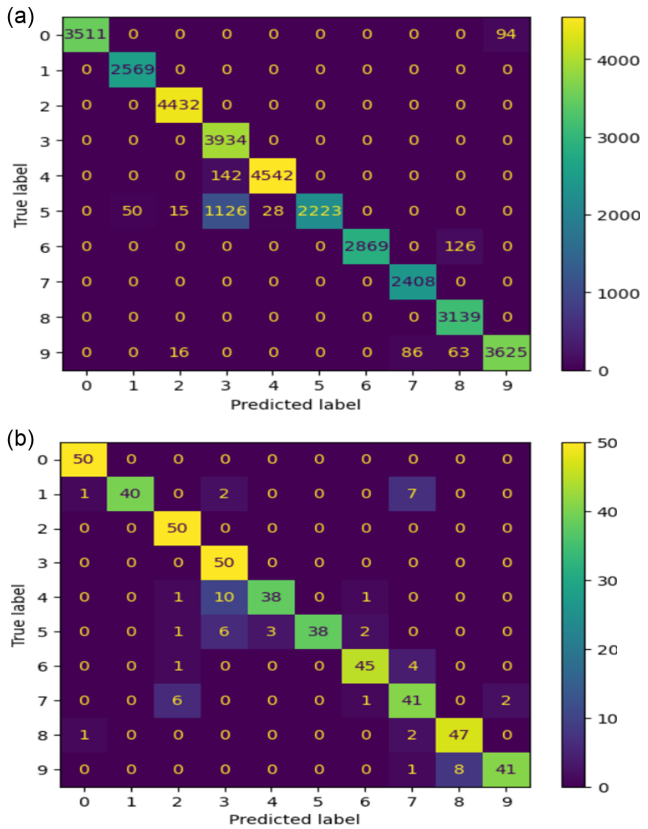


Figure 9
Confusion matrix of the weighted probability ensemble model on (a) MR-AMR main dataset and (b) MR-AMR benchmark-augmented testing dataset/challenge dataset



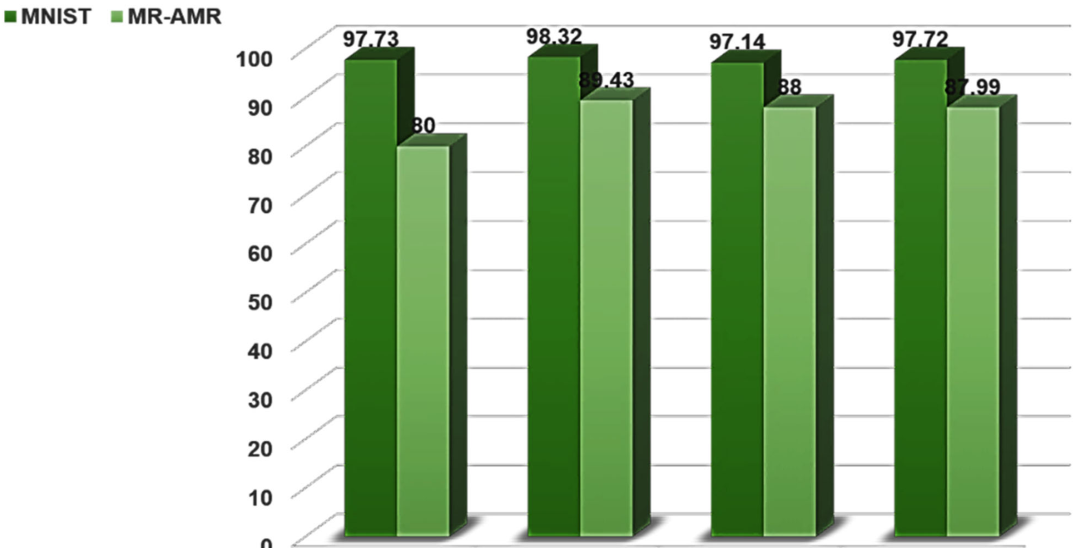
and to show its robustness to the augmentations, we applied to test the model on our benchmarking dataset, as shown in Figure 10.

To validate the performance of the proposed method on MNIST dataset, we conducted experiment, and the results are reported in Figure 10, where the proposed method achieves better results for MNIST dataset than our dataset. The reason is that the MNIST dataset is simple compared to our dataset. Therefore, one can conclude that the proposed work can be used in several real-world applications other than automation of water meter digits recognition.

To show superiority to the existing methods, we implemented [39, 41, 42] state-of-the-art methods to compare with the proposed method. We conducted experiments on our dataset and the results are reported in Figure 11, where it can be noted that the proposed method outperforms the existing methods in terms of accuracy. This make sense because the methods were developed for handwritten digit recognition (MNIST dataset) and the methods do not have generalization ability. On the other hand, the proposed method works well for both MNIST and our dataset (water meter digits recognition).

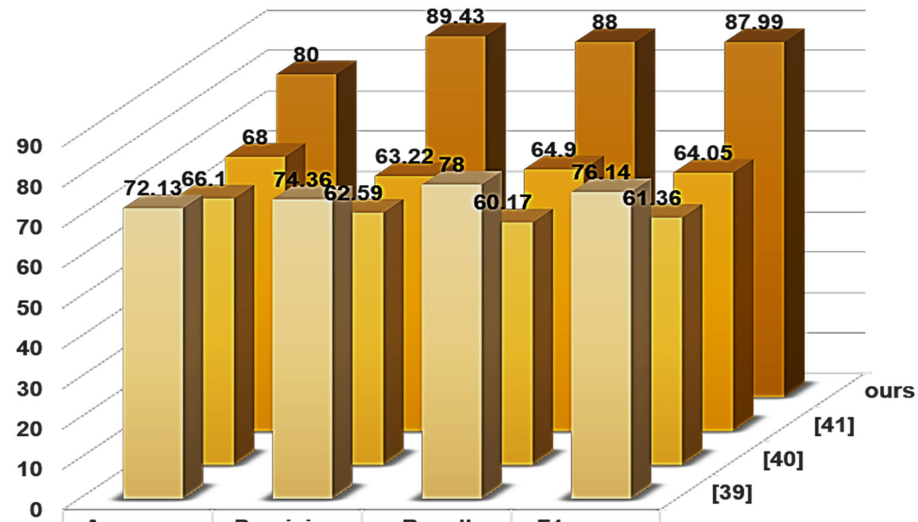
The scope of the method is to recognize the water meter digits for automation. However, the results on MNSIT dataset show (Figure 10) that the proposed method can be extended to handwritten text recognition. This is because our dataset is much more complex compared to MNIST dataset and it includes most the possible handwriting variations. Therefore, the method can be extended to handwritten text recognition. Since the proposed method works well for our and MNIST dataset, one can argue that the proposed method has generalization ability to some extent. This makes sense because unlike existing methods depend on large number of samples and heavy network models, the proposed model is simple and does not depend on predefined samples. In other words, if we fine tune the proposed method with

Figure 10
Confusion matrix of the weighted probability ensemble model on (a) MR-AMR main dataset and (b) MR-AMR benchmark-augmented testing dataset/challenge dataset



	Accuracy	Precision	Recall	F1-score
MNIST	97.73	98.32	97.14	97.72
MR-AMR	80	89.43	88	87.99

Figure 11
Confusion matrix of the weighted probability ensemble model on (a) MR-AMR main dataset and (b) MR-AMR benchmark-augmented testing dataset/challenge dataset



	Accuracy	Precision	Recall	F1-score
[39]	72.13	74.36	78	76.14
[40]	66.1	62.59	60.17	61.36
[41]	68	63.22	64.9	64.05
ours	80	89.43	88	87.99

a few real samples, the proposed method performs well for other datasets like handwritten recognition.

Limitations: It is seen in the confusion matrices (see Figure 9) that our model fails to properly recognize the classes in some cases. There is a noticeable pattern in the misclassifications due to the complexity of the benchmarking dataset and also the main MR-AMR dataset. The similarity in shape and the noise (see Figure 12(a)) causes the misclassifications. Also, another reason for misclassification among classes like 4 and 5, 5 and 6, 6 and 7, etc. is seen in Figure 12(b) representing the complex images of the dataset which are error-prone. Although the proposed method works well for poor quality water meter digit images, sometimes,

Figure 12
Sources of error due to similarity in structures and complexity images. (a) The misclassifications due to structural similarities when noise is added. (b)The complexity of the images, which creates misclassification errors

(a)

Misclassification errors due to noise

(b)

Misclassification errors due to noise

when the images are affected by multiple adverse factors, such as noise, blur, poor contrast, and loss of shapes, the performance of the proposed method degrades. This shows that there is a scope for the improvement.

5. Conclusion and Future Work

We proposed a novel classification ensemble technique utilizing the attention-aided CNN model composed of two base CNNs (ResNet50 and InceptionV3). We conducted a series of experiments, starting with evaluating the individual performance of ResNet50 and InceptionV3 on the MR-AMR dataset, followed by combining the two CNNs, applying FCBAM-aided attention on the combined CNN architecture, and finally, using the weighted probability ensemble technique on the two generated snapshots of the attention-aided combined CNN architecture. Our model achieved an accuracy of 88% on the MR-AMR benchmarking dataset while achieving a high accuracy of 97.73% on the MNIST dataset. The experimental results show that our model outperforms the individual models and provides an effective approach for image classification tasks. We plan to increase the number of benchmarking samples with more complex augmentations to make the dataset more challenging. Also, we plan to design a lightweight CNN model (to make it faster for real-time applications and more efficient) with enhanced performance to reduce the misclassifications that our model faces.

Acknowledgments

We would like to congratulate the entire I2A research team for helping to realize the intelligent onboard system for a mechanical water meter, especially doctoral students who participated in the design of the old MR-AMR dataset and the design of the 3eddad.ma website.

Supportive Materials

The codes of this paper are available at https://github.com/AyushRoy2001/Watermeter-Digit-Recognition.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

Palaiahnakote Shivakumara is an editor-in-chief for *Artificial Intelligence and Applications*, and was not involved in the editorial review or the decision to publish this article. The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The main MR-AMR dataset is available at https://data mendeley.com/datasets/8xjhrrk9rx and the version-2 (which includes the challenge benchmarking dataset) is available at https://www.kaggle.com/datasets/ayush02102001/watermeter-data-recognition.

Author Contribution Statement

Ayush Roy: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Visualization. **P. Shivakumara:** Conceptualization, Validation, Writing – original draft, Supervision, Project administration. **Umapada Pal:** Writing – review & editing.

References

[1] Roy, A., Acharya, A., Biswas, S., Ray, S., & Ganguly, B. (2022). Identification and classification of human mental stress using physiological data: A low-power hybrid approach. In *2022 IEEE 6th International Conference on Condition Assessment Techniques in Electrical Systems*, 136–139. https://doi.org/10.1109/CATCON56237.2022.10077709

[2] Verma, G., Mittal, P., & Farheen, S. (2020). Real time weather prediction system using IoT and machine learning. In *Proceedings of the 2020 International Conference on Smart Cities and Emerging Technologies*, 322–324. https://doi.org/10.1109/ICSC48311.2020.9182766

[3] Fraga-Lamas, P., & Fernández-Caramés, T. M. (2023). Tactical edge IoT in defense and national security. In *IoT for Defense and National Security*, 377–396.

[4] Kanchi, S., Thakkar, P., & Chauhan, D. (2022). EmmDocClassifier: Efficient multimodal document image classifier for scarce data. *Applied Sciences*, *12*(3), 1457.

[5] Mokayed, H., & Mohamed, A. (2014). A robust thresholding technique for a generic structured document classifier using ordinal structure fuzzy logic. *International Journal of Innovative Computing, Information and Control*, *10*(4), 1543–1554.

[6] Khan, M. A. Usman, Abbasi, & Qaiser. (2022). A comprehensive survey of depth completion approaches. *Sensors*, *22*(18), 6969.

[7] Mokayed, H., Ahmed, A., Mustafa, S., & Hamdi, N. (2014). Car plate detection engine based on conventional edge detection technique. In *International Conference on Computer Graphics, Multimedia and Image Processing*.

[8] Veerai-ah, M. C., & Chandrashekara, M. C. (2008). Automated watermeter reading system based on image processing. *International Journal of Computer Science and Network Security*, *8*(5), 67–73.

[9] Wei, Y., Liu, Q., Wei, & Wang, T. (2017). Automatic water meter reading system using machine learning. In *Proceedings 2017 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*, 812–817.

[10] Zhang, J., Li, & Liu, W. (2019a). Deep learning-based water meter reading system. *IEEE Access*, *7*, 9059–9067.

[11] Reddy, B., & Ray, G. G.. (2014). A comparison of digit recognition techniques for automated water meter reading. *International Journal of Engineering and Technology*, *6*(2), 1427–1436.

[12] Rashid, M. M., & Islam, A. T. M. S. (2013). An efficient method for automated water meter reading using edge detection and Hough transform. *International Journal of Computer Applications*, *71*(15), 8–14.

[13] Voon, W. G., Saad, M., Anwar, K., Y., & Kok, B. (2022). Performance analysis of seven convolutional neural networks (CNNs) with transfer learning for invasive ductal carcinoma (IDC) grading in breast histopathological images. *Scientific Reports*, *12*(1), 19200.

[14] Hu, Y., Zhe, Z., Shu, & Hu, H. (2020). Water meter digit recognition using deep convolutional neural networks. *Journal of Physics: Conference Series*, *1529*(1), 012095.

[15] Ghanbari, G., Alimardani, R., Ghodrati, S., M., Khajehzadeh, N., Farsad, S., & Akbari, H. (2019). Automated water meter reading system using ResNet-50 deep learning model. In *Proceedings of the 2019 IEEE Symposium on Computer Applications and Industrial Electronics*, 1–5.

[16] Zhang, H. C., & Liu, W. (2018a). Water meter digit recognition using residual network with region proposal network. In *Proceedings of the 2018 International Conference on Advanced Mechatronic Systems*, 246–250.

[17] Liu, J. & Wang, X. (2018). Water meter digit recognition based on residual networks. In *Proceedings 2018 International Conference on Smart Grid and Clean Energy Technologies*, 446–449.

[18] Wang, J., Zhang, & Liu, Y. (2019a). Water meter reading system using residual networks and transfer learning. In *Proceedings 2019 International Conference on Intelligent Transportation, Big Data & Smart City*, 161–165.

[19] Park, S., Lee, K., Min, & Nam, I. K. (2020). Deep learning-based water meter reading using Inception-ResNet-v2 architecture. *Sensors*, *20*(10), 2986.

[20] Zheng, Z., & Zhang, J. (2018). Water meter reading based on Inception-v3 network and data augmentation. In *Proceedings 2018 IEEE 5th International Conference on Cloud Computing and Intelligence Systems,* 536–539.

[21] Zhang, Q., Wang, L., Chen, X., & Liu, Y. (2018b). Water meter reading with Inception-v4 network. In *Proceedings 2018 IEEE 3rd International Conference on Big Data Analysis*, 90–94.

[22] Zhou, J., Liu, W., Zhang, & Li, M. (2019). Water meter reading using Inception-v3 network and deep convolutional neural network. In *Proceedings of the 2019 3rd International Conference on Electrical, Control and Automation Engineering*, 128–132.

[23] Guo, S. (2018). Water meter reading system based on Inception-v4 network with data augmentation. In *Proceedings 2018 International Conference on Smart Grid and Clean Energy Technologies*, 300–303.

[24] Yan, J. H., & Yang, W. (2021). Water meter reading with combined Inception-ResNet-v2 network. In *IEEE 2021 11th International Conference on Computer and Communications Management*, 68–73.

[25] Liu, X. (2021). Water meter reading based on combined ResNet and Inception network. In *Proceedings 2021 IEEE 2nd International Conference on Artificial Intelligence and Pattern Recognition*, 293–298.

[26] Zhang, W., Li, & Xu, H. (2019b). Water meter reading with combined Inception-v4 and ResNet50 network. In *IEEE 2019 International Conference on Smart Grid and Clean Energy Technologies*, 425–429.

[27] Sun, W., Li, M., & Wang, Z. (2021). Water meter reading with combined inception and ResNet network using ensemble learning. In *2021 IEEE 2nd International Conference on Artificial Intelligence and Pattern Recognition,* 289–292.

[28] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *arXiv Preprint: 1807.06521*.

[29] Luo, Z., Xi, H., Zhi, C., & Li, X. (2019). Dual attention networks for multi-spectral pedestrian detection. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 5826–5835.

[30] Nakanishi, R., & Matsumoto, Y. (2021). CBAM-based deep learning model for digit recognition. *Applied Sciences*, *11*(8), 3604. https://doi.org/10.3390/app11083604

[31] Zhang, W., Liu, Y., & Zhang, X. (2021). Attention-based capsule network for handwritten digit recognition. *Journal of Applied Remote Sensing*, *15*(1), 16512. https://doi.org/10.1117/1.JRS.15.016512

[32] Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J. E., & Weinberger, K. Q. (2017). Snapshot ensembles: Train 1, get m for free. *arXiv Preprint: 1704.00109*.

[33] DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *arXiv Preprint: 1708.04552*.

[34] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

[35] Li, X., Liu, S., & Zhang, L. (2016). Automatic water meter reading system based on image processing. In *2016 IEEE International Conference on Mechatronics and Automation*, 1037–1042.

[36] Zhang, T., Wang, X., Zhao, W., & Liu, J. (2018c). Real-time water meter reading using deep learning. In *2018 IEEE International Conference on Mechatronics and Automation*, 468–473.

[37] Wang, T., Liu, Q., & Li, H. (2019b). A comparative study of deep learning methods for water meter reading. In *2019 IEEE 4th International Conference on Computer and Communication Systems*, 205–210.

[38] Chen, G., Wang, L., Zhang, X., & Liu, W. (2011). Water meter digit recognition using histogram-based features and support vector machines. In *2011 IEEE International Conference on Computer Science and Automation Engineering*, 144–148.

[39] Keerthi, T. (2022). MNIST handwritten digit recognition using machine learning. In *2022 IEEE 2nd International Conference on Advance Computing and Innovative Technologies in Engineering*, 768–772.

[40] Naim, A., Aaroud, A., Akodadi, K., & El Hachimi, C. (2021). A fully AI-based system to automate water meter data collection in Morocco Country. *Array*, *10*, 100056. https://doi.org/10.1016/j.array.2021.100056

[41] Masthan, P., Rao, & Patel, A. (2023). MNIST digital classification and handwritten digit recognition. *International Journal of Scientific Research and Engineering Trends*, *9*(2), 279–282.

[42] Siddhartha, P. V. (2023). Digit recognition of MNIST handwritten using convolutional neural networks (CNN). In *2023 IEEE International Conference on Intelligent Systems for Communication, IoT and Security*, 328–332.

How to Cite: Alkhaled, L., Roy, A., & Shivakumara, P. (2023). An Attention-Based Fusion of ResNet50 and InceptionV3 Model for Water Meter Digit Recognition. *Artificial Intelligence and Applications* https://doi.org/10.47852/bonviewAIA32021197