Water Meter Number Recognition using Multilayer Convolutional Neural Network

U COM 6

Abstract. Water Meters are used to calculate water consumption. The meter measures the volume of water used by the facility and displays it. To find the amount of numbers used, a human will have to read the number and record it. However, this is both time consuming and prone to human error. This study aims to build a multilayer Convolutional Neural Network (CNN) to recognize multi-digit numbers in real-world water meter images. For the dataset, five thousand water meter images were collected to build three water meter image datasets, a single digit dataset from cropped five-digit water meter images, a five-digit dataset from cropped water meter images and an original water meter image dataset. The research found that the CNN model is trained on the single digit dataset and achieved with 98% accuracy. Additionally, the CNN model is trained on the five- digit dataset and achieved with ~85% accuracy. The CNN model is also trained on the water meter image dataset and achieved with ~45% accuracy. The project uses the TensorFlow library in Python. This learning model and method can be applied to other forms of number recognition. The data can be used to help with models that use non-five digit number strings or flexible number strings.

Literature review

Neural Network Modeling

The convoluted neural network (CNN) is one of the most widely used frameworks to identify and recognize images. A CNN is used to examine and recognize features while a recurrent neural network (RNN) is used to create the model. The RNN isn't as efficient or quick as a CNN. Speed and efficiency are especially important in machine learning models as the entire point is to have it be able to be faster than a human. Though in some cases the goal is to just be able to replace a human. In a study by the South China University of Technology (Yang & Jin 2019), a dataset called SCUT-WMN was constructed for the purpose of open research. The

FCSRN was proposed for fast and accurate research. Many layers of additional neural networks and variables were added. It was concluded that the FCSRN did not need recurrent layers, and only needed less parameters and computer power. The FCSRN is shown to be more efficient than RNN models.

Water Meter Readings

A study at the Chinese Academy of Sciences (Gao 2017) attempted to find a way to make water meter readings more efficient by the use of AI on mobile devices. As it was time consuming and labor intensive to record the water gauge and write them down by hand, it was determined that taking a photo of the gauge and having the AI read and record it would be much better. Water meter recognition AI have either used template matching or BP neural networks. A new approach was created and had the following steps. Digit detection, digit recognition, sequence feature, and attention decoder. ACF features were boosted to facilitate number detections. This new approach mainly consisting of digit detection and recognition found that detection was objectively much more efficient. Another new find is that this technique can be utilized without separating segments of the digits of the water meter. This improves efficiency and accuracy greatly. Adding Spatial Transform Network to the model is in their future plans.

A study by (Li Su Yuan and Chu 2019) focused upon the CNN based automatic water meter reading in a smart city. This study differed from the norms of using OpenCV. Instead, a novel light-weighted spliced convolution network was used to recognize the digits of the water meter number. Convolutions were 3x3, kernels were 1x1 and 3x3. The model was utilized on real world datasets and found to be more efficient. The model in this study could reduce computational consumption by nearly ten times, storage by seven, and run time by 3. Future plans include other perspectives of the photo (camera in different positions).

Convoluted Neural Network

A study at CMU (Lawrence 1997) looked at a hybrid neural-network model at identifying faces. The model consists of local image sampling, a self-organizing map neural network, and a CNN. The self-organizing model increases efficiency, however, the model itself takes a long time to train. On the other hand, CNN also takes a long time to train.

Another ability of machine learning is to identify and recognize the language, sentences, and meaning. A study at Oxford (Kalchbrenner et al 2014) looked at a dynamic CNN and its language identification. The model used dynamic k-max pooling. A point of interest of the model is that it can be applied to all languages. Four experiments were run: small scale binary, multi-class sentiment prediction, six-way question classification, and twitter sentiment prediction. The model did well with the first 3 tasks. The model had the following abilities: sentiment analysis, paraphrase detection, entailment recognition, summarisation, discourse analysis, machine translation, grounded language learning, and image retrieval.

The Harbin Institute of Technology also looked at CNN and its uses on language learning(Hu, Lu, Li, Chen 2014). A CNN model tasked with matching two sentences was proposed to fill the objective of modelling the internal structures between language objects and their interactions.

Deep Convoluted Neural Network

The University of Toronto (Krizhevsky et al 2012) conducted research to gauge the limits of convolutional neural networks. Usual datasets for machine learning models range to the thousands on average. The dataset utilized here had 1.2 million high-resolution images with 1000 total classes. The dataset had a top1 error rate of 37.5% and top5 error rate of 17%. The study concluded that the limits of a CNN was record-breakingly high. It can achieve extremely high numbers with just supervised learning. The deep model dropped in performance if even a single convolutional layer was removed. As the network becomes larger, results will improve. Future study will include deep learning on video sequences.

(Xu et al, 2014) led a study that investigated deep CNN. Instead of focusing on their mathematical or physical characteristics, they use previous samples to train the model. This requires a large amount of data. The task of the model was image deconvolution.

Shufflenet

Shufflenet is an extremely efficient CNN architecture introduced by (Zhang, Zhou, Lin, Sun 2018). It is designed specifically for mobile devices which have very limited computer power. Two variables of interest are pointwise group convolution and channel shuffling. A new operation "channel shuffle" was created to help information flow across channels.

Super-Resolution Convolutional Neural Network

The process of restoring a low-image quality photo to one of a higher quality was examined by (Dong, Loy, Tang 2016). SRCNN has demonstrated its proficiency in processing, but has the detriment of requiring a large amount of computational power. The study redesigned the SRCNN structure to require less computational power. The goal is less power yet still able to render at 24 fps. The three changes include: addition of a deconvolutional layer, reformation of a mapping layer, and smaller filter sizes but more mapping layers. SRCNN is already really proficient but this study has improved upon it by getting rid of its weakness. It is now able to be adapted to real-time uses because it can render at 24 frames per second.

Discussion

The super-resolution convolutional neural network can prove to be very helpful in any time of machine learning using images. Especially when taking photos of water meters, it is quite likely that a few photos may be blurred. The SRCNN model can be included to attempt to de-blur an image so that it can be used for image learning, though this might decrease efficiency.

Facial recognition models are similar to number recognition as they are both machine learning based on similarities on photos. The difference is that number recognition is the identification of one image and sorting it into one of many classes. Facial recognition is the opposite.

Shufflenet can prove to be useful if the machine learning task is to be carried out on the phone for efficiency purposes. Though it is more likely that the phone will be used to take photos and the number recognition will be done on a computer. Though in the future, if a company wants maximum efficiency, they will go down the shufflenet route.

The deep convolutional neural network will be helpful in water meter recognition because the dataset available is extremely large. The dataset is essentially every single water meter with a visual interface.

The current water meter machine learning tasks have shown what has been studied and what has not been. Therecan be many differences in the way the model is processed that may impact efficiency. Whether the digits are segmented, whether the images are cropped, whether the image is transformed to black and white and many other variables. It will be helpful for future studies on this topic if a study examining the efficiency of brute force machine learning is done. Though it hasn't been done yet because most studies like to examine how to improve efficiency.

Rationale. Water Meters are used to calculate water consumption. The meter measures the volume of water used by the facility and displays it. It can measure total volume used or even the volume of a portion of the system. To find the amount of numbers used, a human will have to read the number and record it. However, this is both time consuming and prone to human error. Newer automatic electronic readers are too expensive to implement in some poorer areas. This study aims to build a machine learning model using a convoluted neural network that will read and interpret the digits of a water meter gauge to facilitate the recording effort.

Hypothesis: It is not possible to create a model that can read and interpret water meter readings

Null Hypothesis: It is possible to create a model that can read and interpret water meter readings to a high efficiency.

Independent Variable: Build of network model, options used to create model, optimization choices, dataset source and size.

Dependent Variable: Accuracy of recognition model.

Control Variable: Coding language and package

Procedures/Data Analysis: Convoluted Neural Network (CNN) modeling will be used to develop/train a machine learning model. For the dataset, five thousand water meter images were collected to build three water meter image datasets: a single digit dataset from cropped five-digit water meter images, a five-digit dataset from cropped water meter images and an original water meter image dataset. These datasets will be taken from the five thousand water meter images source and cropped to only show the number gauge. The digits will be noted in a spreadsheet with the corresponding image file name.

Risk and Safety: None

Results. The model trained from dataset Water Model 3 (WM3) contained three convolutional layers and 2 fully connected layers (Table 1). It was trained on 4800 training and 1200 validation images and achieved a 98% accuracy. Single digit images have relatively simple image structures which are captured by the model with high accuracy. For five digit number recognition, the model with five outputs as shown in Table 2 and Figure 4 was trained on 3500 training and 800 validation images using images from dataset Water Meter 2 (WM2) and achieved 85% accuracy. Table 1 and 2 represent the settings of the model and what options were used trying to maximize accuracy of the model. Figure 4 is a visual interpretation of how the network model was built. One way to achieve better accuracy is to acquire more training images. The model trained from dataset Water Meter 1 was trained on 3500 training and 800 validation images. The model achieved 45% accuracy. The accuracy differences between dataset WM1 and dataset WM2 reflect that the model cannot fully capture multiple digit number image structures due to low true signal (e.g. number of training images) to noise ratio (e.g. images irrelevant to digit number images). Images irrelevant to digit number images refer to images that were taken but had illegible numbers or had no visible water meter.

Layer/Operation	Parameters	Output Size
Inputs	-	100, 100, 1
Convolutional Layer	3x3 kernel	98, 98, 64
Max Pooling	2x2 filter	49, 49, 64
Convolutional Layer	3x3 kernel	47, 47, 64
Max Pooling	2x2 filter	23, 23, 64
Convolutional Layer	3x3 kernel	21, 21, 128
Max Pooling	2x2 filter	10, 10, 128
Convolutional Layer	3x3 kernel	8, 8, 128
Max Pooling	2x2 filter	4, 4, 128
dropout	0.5	2048
Fully Connected Layer	ReLU activation	512
Fully Connected Layer 1	Softmax activation	10

Table 1. Network Architecture I

Layer/Operation	Parameters	Output Size
Inputs	-	100, 100, 1
Convolutional Layer	3x3 kernel	98, 98, 64
Max Pooling	2x2 filter	49, 49, 64
Convolutional Layer	3x3 kernel	47, 47, 64
Max Pooling	2x2 filter	23, 23, 64
Convolutional Layer	3x3 kernel	21, 21, 128
Max Pooling	2x2 filter	10, 10, 128
Convolutional Layer	3x3 kernel	8, 8, 128
Max Pooling	2x2 filter	4, 4, 128
dropout	0.5	2048
Fully Connected Layer	ReLU activation	512
Fully Connected Layer 1	Softmax activation	10
Fully Connected Layer 2	Softmax activation	10
Fully Connected Layer 3	Softmax activation	10
Fully Connected Layer 4	Softmax activation	10
Fully Connected Layer 5	Softmax activation	10

Table 2. Network Architecture II

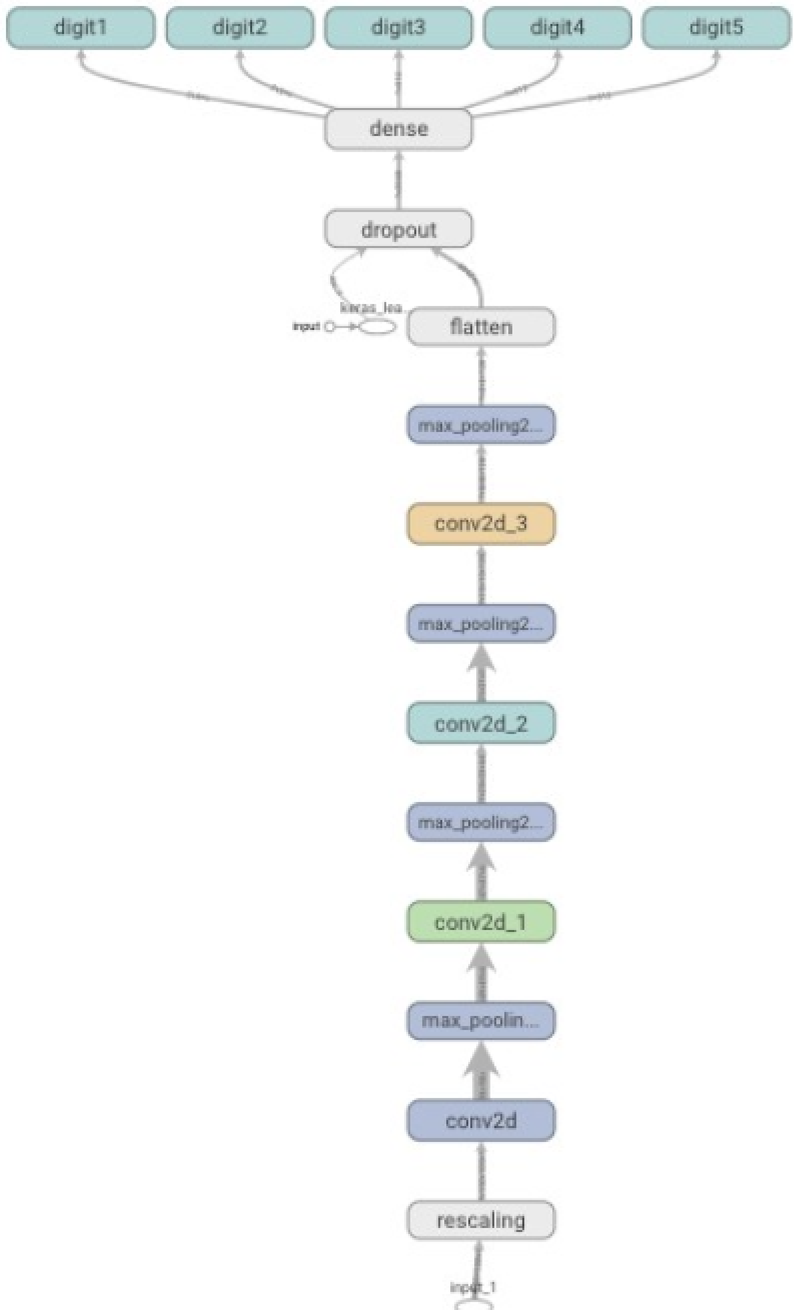

Figure 4, Architecture of Entire Network

Discussion. This study utilizes a multi-layer CNN model that recognizes numbers from images. A CNN model is a class of deep neural networks most commonly used in image recognition. CNN model was chosen because it is seen as advantageous in its image recognition ability as opposed to recurrent neural networks. 5000 real world water meter pictures were used for the datasets. One example is shown in figure 1. They were manually cropped to look like figure 2, and then separated and identified by their value like figure 3 using python.

Figure 1 Water Meter Image

Figure 2 Cropped Five Digit Numbers

Figure 3 Cropped Single Digit Number

In an attempt to raise accuracy, transfer learning using the SVHN dataset has been attempted. Through fine-tuning on pre-trained models, it did not show significantly improved accuracy on the dataset WM2 and WM3. The SVHN is a commonly used real-world image dataset provided by Stanford University.

Object detection with neural networks can be used to localize individual objects on images. However, the object detection model only localizes several digit numbers and fails to localize multiple digit numbers as whole on water meter images.

Conclusion. It is demonstrated that a multi-layer convolutional neural network can be used to recognize multiple digit numbers on water meter images with image single number cropping. Multiple digit numbers on water meter images can also be recognized simultaneously with the multi-layer convolutional neural network. This study and the models created strengthen image recognition in general as this study will be contributed to the artificial intelligence community and can be used in other research in the future.

The massive dataset manually created is not only useful to this study but also future studies anywhere as well. A larger pool of data is always useful as it can raise model accuracy. This dataset can be applied to not only other water meter image recognition, but also real-world number recognition that does not pertain to only water meters. The individually cropped numbers can be applied to any dataset that looks at individual single unit numbers. As such, there are well over 25000 images from this dataset itself. These images can be given upon request by the author.

Bibliography

Y. Lecun and C. Cortes. The MNIST database of handwritten digits,

http://yann. lecun. com/exdb/mnist/, 1998

Sermanet, P., Chintala, S., & LeCun, Y. (2012). Convolutional neural networks applied to house numbers digit classification. In ICPR 2012 - 21st International Conference on Pattern Recognition (pp. 3288-3291). [6460867]

Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet (2014). Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks. arXiv:1312.6082v4.

VGG Image Annotator, web site at

http://www.robots.ox.ac.uk/~vgg/software/via/

Yang, F., Jin, L., Lai, S., Gao, X., & Li, Z. (2019). Fully convolutional sequence recognition network for water meter number reading. *IEEE Access*, *7*, 11679-11687.

Gao, Y., Zhao, C., Wang, J., & Lu, H. (2017, August). Automatic watermeter digit recognition on mobile devices. In *International Conference on Internet Multimedia Computing and Service* (pp. 87-95). Springer, Singapore.

Li, C., Su, Y., Yuan, R., Chu, D., & Zhu, J. (2019). Light-weight spliced convolution network-based automatic water meter reading in smart city. *IEEE Access*, *7*, 174359-174367.

Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, *8*(1), 98-113.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6848-6856).

Dong, C., Loy, C. C., & Tang, X. (2016, October). Accelerating the super-resolution convolutional neural network. In *European conference on computer vision* (pp. 391-407). Springer, Cham.

Xu, L., Ren, J. S., Liu, C., & Jia, J. (2014). Deep convolutional neural network for image deconvolution. In *Advances in neural information processing systems* (pp. 1790-1798).

Hu, B., Lu, Z., Li, H., & Chen, Q. (2014). Convolutional neural network architectures for matching natural language sentences. In *Advances in neural information processing systems* (pp. 2042-2050).