IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 1, 2019, accepted November 19, 2019, date of publication November 28, 2019, date of current version December 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956556

Light-Weight Spliced Convolution Network-Based Automatic Water Meter Reading in Smart City

CHUNSHAN LI[1], YUKUN SU[2], RUI YUAN[3], DIANHUI CHU[1], AND JINHUI ZHU[2]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Weihai 264209, China
[2]School of Software Engineering, South China University of Technology, Guangzhou 510640, China
[3]School of Journalism and Communication, The Chinese University of Hong Kong, Hong Kong

Corresponding authors: Dianhui Chu (chudh@hit.edu.cn) and Jinhui Zhu (csjhzhu@scut.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0830900, in part by the Science and Technology Planning Project of Guangdong, China, under Grant 2017B090910005, Grant 2018B010108002, and Grant 2015A030310446, in part by the National Natural Science Foundation of China under Grant 61876208, Grant 61772159, and Grant 61902090, in part by the Pearl River S&T Nova Program of Guangzhou under Grant 201806010081, in part by the Pre-Research Foundation of China under Grant 61400010205, in part by the Foshan Scientific and Technological Innovation Funds under Grant zk2018007, and in part by the Natural Science Foundation of Shandong Province under Grant ZR2017MF026.

ABSTRACT Automatic reading for water meter is one of the practical demands in smart city applications. Due to the high cost, it is not feasible to replace the old mechanical water meter with a new embedded electronic device. Recently, image recognition based meter reading methods have become research hotspots. However, illumination, occlusion, energy and computational consuming in IoT environment bring challenges to these methods. In this paper, we design and implement a smart water meter reading system to handle this issue. Specifically, we first propose a novel light-weight spliced convolution network to recognize the meter number, which simplifies standard 3×3 convolutions by splicing a certain number of 1×1 and 3×3 size kernel. We then prove the superiority of our network by theoretical analysis. Second, we have implemented the prototype which can handle huge real-time data base on the distributed cloud platform. Base on this system, our system can provide industrial service. Finally, we conduct real-world dataset to verify the performance of the system. The experimental results demonstrate that our proposed light-weight spliced convolution network can reduce nearly $10\times$ computational consuming, $7\times$ model space, and save $3\times$ running time comparing with standard convolution network.

INDEX TERMS Water meter reading, spliced convolution network, cyber-physical-system, smart city.

I. INTRODUCTION

The world is increasingly looking forward to adapting and using new technologies to improve the quality of life and to reduce the environmental impact of human activities and consumption patterns [1]. Automatic reading for water meter is one of the practical demands in smart city applications, by which the government can monitor water consumption in real-time and water companies can significantly save labor costs. However, old water meter without data upload function is the most common device in water transmission systems over several decades, and it cannot be replaced by the electronic meter in the foreseeable future. Among the water meter number reading methods proposed in the literature [2], the means of reading of water meter can be summarized into manual transcription, sensors based methods, image recognition technology-based methods.

Manual transcription employs humans to read and record water meter numbers, which are labor cost wasting. In particular, one meter-reader will estimates hundreds and thousands of meters per-day. It often happens that meter-reader records a wrong number which can cause economic losses and disputes to users and water companies. Adding a data-uploading module to the water meter is important progress for a a water meter reading. However, the replacement process needs to disassemble and install the water meter. The massive old devices in urban make the project more troublesome and laborious [3]. In recent years, image recognition technology-based methods have begun to research hotspots. Relative to the installation of the independent module inside the water meter, these methods only need to place a camera on the surface of the water meters. And then, these methods will photo the meter images,

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Yu.

recognizes the numbers in the images, and transfer these records to cloud service [4]. Obviously, these methods are highly flexible, easy to implement, and material resources cost is relatively low.

There are two kinds of solutions for image recognition technology-based methods. The first category is OpenCV image recognition technologies [5], [6]. These methods have been able to identify some objects such as numbers, image object contours, etc., and can accomplish simple, practical tasks on image processing tasks. However, in the face of industrial applications such as water meter recognition, strong light, low light, occlusion, incompleteness, and other external factors often bring the poor effect. Another category to recognize meter number is deep neural network (DNN) based methods. Early DNN based technologies called segmentation-based methods [7], [8], which always include two steps: character segmentation and recognition, e.g., Bissacco *et al.* [8] first use a bounding box to fix the position of number in image and then employ a convolutional neural network (CNN) to recognition objective. Recent DNN based methods are called the segmentation-free methods [9], [10], which utilize recurrent neural network (RNN) to recognizes the number in an image as a whole without fixing the position of objects. Although the DNN based methods achieve good performance in a number recognition task, these still exist several challenges to hinder their applications in industrial environments. For example, a water company starts the tasks of meter reading on a monthly period. On the starting point, huge data will flow into the recognition system, and which inspires us to design more flexible methods with high speed and low computational consuming.

In order to solve the robustness and computational consuming problems, in this paper, we design and implement smart water meter reading system base on the distributed cloud platform. First, we propose a novel light-weight spliced convolution network to recognize the meter number, which simplifies the number of 3×3 convolutions by splicing a certain number of 1×1 and 3×3 size kernel to reduce the entire convolution parameters and network computational consuming. Therefore, our model can handle large amounts of meter images in the terminal within high-speed. Second, we built an industrial prototype based on distributed platform, in which the data perception layer uses a camera to capture the reading number on the water meter. The transport layer employs several communication devices (WIFI or 4G) to transfer the data to the function layer for reading recognition. The function layer of the system is based on a cloud platform, on which the calculation and storage are performed via distributed mode. Finally, we conduct real-world dataset to verify the performance of the system. The experimental results demonstrate that our proposed light-weight spliced convolution network can reduce nearly $10\times$ computational consuming, $7\times$ model space, and save $3\times$ running time, compared with standard convolution kernel. The results also prove that our system is a feasible industrial application.

In summary, our main contributions in this paper are stated as follow:

- We propose robustness and high-speed spliced convolution network to recognize the meter number, which simplifies standard 3×3 convolutions by splicing a certain number of 1×1 and 3×3 size kernel. We then prove the superiority of the proposed network model by theoretical analysis.
- We built a prototype for automatic water meter reading in smart city, which can handle huge real-time meter images base on the distributed cloud platform.

The remainder of this paper is organized as follows. Section 2 gives a brief review of the related work. Section 3 details the light-weight spliced convolution network model. Section 4 introduces the the experiments and results. Section 5 shows the experiments and results. And Section 6 conclude the paper.

II. RELATED WORK

As a practical demand in smart city applications, the automatic water meter reading system is growing day by day [11]–[14]. There exist two solutions to read the number of the water meter. The first method to measure the water flow rate is designing and implementing an embedded water meter. In literature [15], authors obtain water flow information automatically by sensing the magnetic flux lines generated by the internal rotating coupling magnets of the water meter. Literature [16] implements an automated meter reader device that has an optical sensor with automatic gain control. Literature [17] presented an automatic meter reading system using encoding technology, which is more easily to industrialization. However, there are a large number of old mechanical water meters in the current urban. The installation of new embedded meter requires re-update or re-installation of the old devices, resulting in the consumption of a large amount of human and material resources. The massive old devices make the replacement project infeasible.

The second method is using a camera to capture water meter images and then employ image recognition technology to read the meter number. Literature [18] proposed an effective method for discriminating secondary recognition based on digital character features. Literature [19] accomplished water meter automatic recognition by using a neural network for numbers in the image. However, the above approaches are usually based on character line detection. When the distance between numbers is large, these approaches achieve poor performance. In order to overcome the assumption of character line, researchers proposed image semantic segmentation based approaches [20], [21]. By semantic segmentation technology, images can be characterized as different classifications of pixel points. With these features, different numbers can be classified easily. Long *et al.* [22] proposed FCN net, which use full convolutional network to extract features for pixel2pixel image segmentation. On this basis of [22], BiSeNet [21] constructed two channels model, where the spatial path is responsible for extracting shallow information, and

FIGURE 1. **Roller type digital water meters.**

the context path is responsible for extracting deep semantic information. ICNet [20] mainly divided the segmentation task into multiple resolutions. In ICNet, the low-resolution output is sent to a high-level resolution network for feature fusion before the classification layer. Literature [23] presented a deep convolutional neural network (CNN), which can raise image processing to a higher level.

Considering the conversion to the digital symbol of the water meter reading requires multiple additional steps, the image semantic segmentation based approach is not the optimal choice. At present, the object detection model has attracted the attention of researchers, which can recognize the meter number in one step. Redmon *et al.* [24] proposed a fast end-to-end detection method based on global information of images for prediction. Liu *et al.* [25] proposed an image multi-scale fusion method, which can effectively solve the detection of objects of various sizes in the image. RFBNet [26] presented a novel receptive field module that makes the detection model faster and more accurate by using multi-branch pools with different convolution kernels corresponding to different sizes of RF. Liu et al. use NMS [27] to filter the target first and then classify the number from 0 to 9. In the industrial area, Google's open-source OCR algorithm tesseractocr [28] and Baidu AI open platform digital identification interface can be used for the recognition of water meter readings.

III. METHOD

In this section, we will introduce the core method of the smart water meter reading system. We will first give the description and challenges for the water meter image reading. And then, the details of the light-weight spliced convolution network will be given.

A. WATER METER IMAGE AND ITS CHALLENGES

For the study of water meter reading, we constructed an image dataset, which is captured by a camera in the real-world environment. Figure 1 shows four old-fashioned water meter with a roller type reading. In this article, all the solutions we want to discuss and solve are based on the old roller-type readings. The challenges and difficulties we have to deal with can be summarized as follows:

- The surface of the water meter is photographed by the camera, and there are an external light, strong light, weak light, and even angles, which will cause potential interference to subsequent work.
- In the digital area of the surface of the water meter, there will also be the influence of other blocks in the image, such as water meter model number, power number, and other text on the water meter.
- Due to the principle of mechanical construction, when one-meter number needs to be read as shown in the picture at the bottom of Fig. 1, there is a gap and two numbers in the same area, which will plague the calculation model to identify, leading to misjudgment or missed judgment.

Based on the above situation, we collected data from various scenarios and added disturbances to increase normalization ability.

B. LIGHT-WEIGHT SPLICED CONVOLUTION NETWORK

This section will introduce details of light-weight spliced convolution network, including backbone network, spliced convolution, and disordered combination splicing. As shown in Figure 2, the architecture of the meter number reading model is end-to-end. The network architecture of our proposed method consists of different convolution stages, which are explicitly extracting features and downsampling, and the operation in each stage is performed by the proposed spliced convolution module.

1) BACKBONE NETWORK

We use a network in [29] as the backbone network, in which a standard convolutional layer takes as input a $D_H \times D_W \times D_M$ feature map $\mathbf{F}$ and produces a $D_H \times D_W \times D_N$ feature map $\mathbf{G}$ where D_H and D_W is the spatial height and width of a square input feature map, D_M is the number of input channels (input depth), D_G is the spatial width and height of a square output feature map, and D_N is the number of output channel (output depth).

The standard convolutional layer is parameterized by convolution kernel $\mathbf{K}$ of size $D_K \times D_K \times D_M \times D_N$ where D_K is the spatial dimension of the kernel assumed to be square, D_M is number of input channels and D_N is the number of output channels as defined previously.

The output feature map for standard convolution assuming stride one and padding is computed as:

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,m} \tag{1}$$

Then there is, the calculation of the standard convolution is as follows:

$$F_{L_S} = D_H \cdot D_W \cdot D_K \cdot D_K \cdot D_M \cdot D_N \tag{2}$$

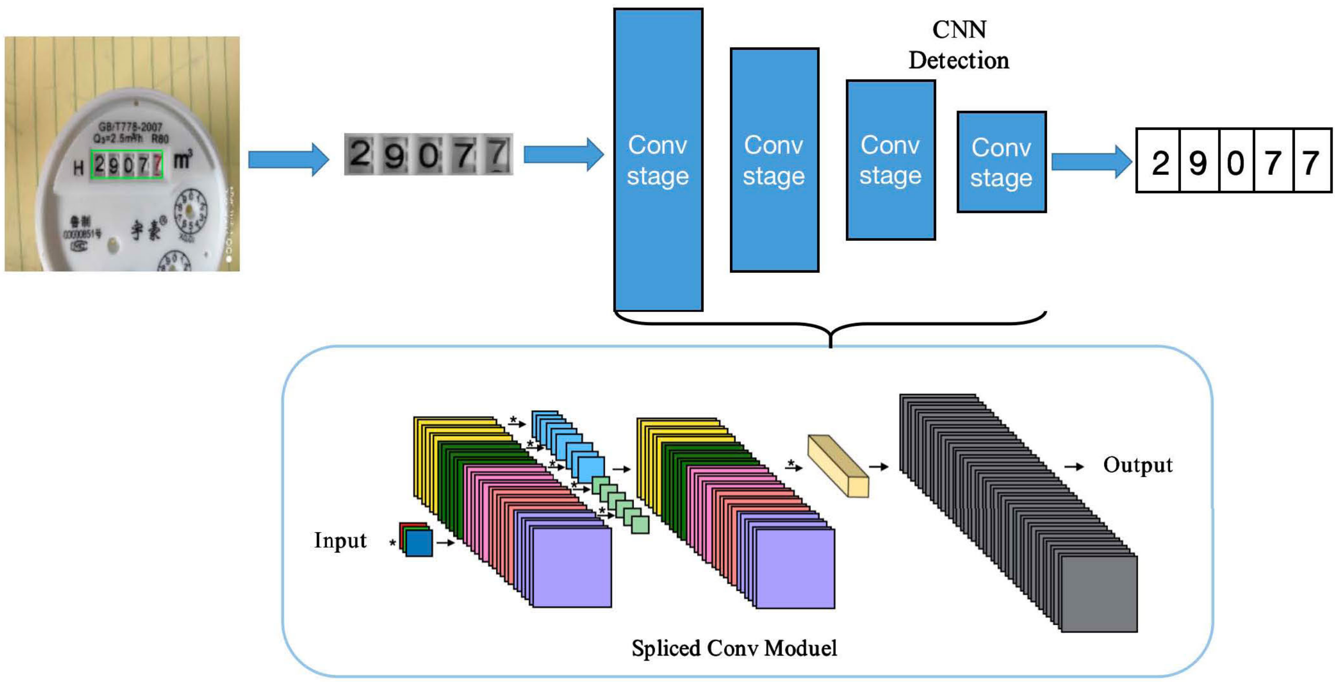


FIGURE 2. The proposed network architecture for water meter recognition.

However, in depthwise separable convolution, the m^{th} channel of the input acts on the m^{th} deep convolution kernel, producing an output feature map. Then immediately follow the 1×1 convolution kernel to raise the dimension. Therefore, the calculation method of the standard convolution is available, and the operation of the depth separable convolution is performed as follows:

$$F_{L_K} = D_H \cdot D_W \cdot D_K \cdot D_K \cdot D_M \tag{3}$$
$$F_{L_1} = D_H \cdot D_W \cdot D_M \cdot D_N \tag{4}$$

The calculation formula reduction (R) by comparing the depthwise separable convolution and traditional standard convolution is as follows:

$$R_{DwConv/Conv} = \frac{F_{L_K} + F_{L_1}}{F_{L_S}} = \frac{1}{D_N} + \frac{1}{D_K^2} < 1 \tag{5}$$

It can be seen that by using this convolution method, the floating-point operation can be reduced, and the overall network parameters are reduced.

In the convolution network, the overall objective loss function is a weighted sum of the localization loss (loc) and the confidence loss (conf) in recognition, which means that there is one more branch regression function than the image classification. In [25], researchers let $x_{ij}^p = \{1, 0\}$ be an indicator for matching the i-th default box to the j-th ground truth box of category p. Following the matching strategy above, we can have $\sum_i x_{ij}^p \geq 1$. The loss function is performed as:

$$L_{x,c,l,g} = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \tag{6}$$

where N is the number of matched default boxes, if N = 0, wet set the loss to 0. The localization loss is a Smooth L1 loss [30] between the predicted box (l) and the ground truth box (g) parameters.

2) SPLICED CONVOLUTION

It is well known that choosing the proper convolution kernel can greatly improve the efficiency of the CNN algorithms. Using depthwise separable convolutions and group convolution can reasonably reduce the computational effort of 3×3 convolutions [31], [32]. In this article, we design a novel convolution operation, called splice convolution, which can use different convolution kernels size such as 1×1 and 3×3 convolution kernels together to form an independent convolution kernel. The advantage of this kernel is that it further reduces the amount of computational complexity caused by the 3×3 convolution kernel, and is not completely calculated by the 1×1 convolution kernel so that a certain amount of information will not be lost. Figure 3 shows the key difference between the splice convolution and other convolutions. The standard 3×3 convolution kernel is calculated by adding the convolution kernel of each dimension to the pixels on the feature map, and the dimensions of the output depend on how many convolution kernels. The depthwise separable convolution is that the different dimensions of the convolution kernel are responsible for the different dimensions of the feature map, and the final output dimension is equal to the input dimension. In splice convolution, we have constructed a light-weight module. In this module, when our

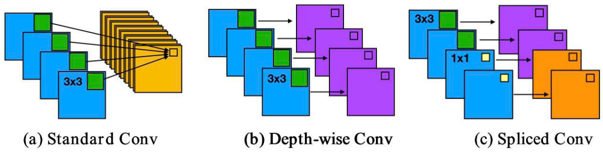


FIGURE 3. Different convolution kernels.

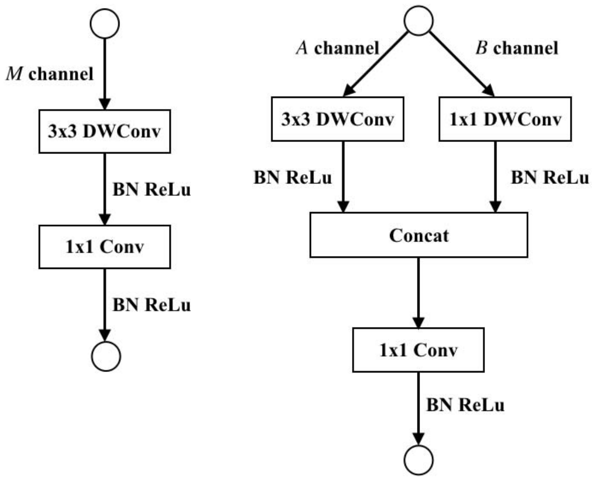


FIGURE 4. Left: Depthwise separable convolution layer with batch- norm and ReLU. Right: Splice convolution layer with batchnorm, ReLU and concat.

splice convolution is operated with the feature map, each dimension of the feature map is grouped separately. Some of them act on 3×3 size kernel, and some act on 1×1 size kernel. Therefore, the proposed splice convolution further reduces the entire convolution kernel and network calculation parameters.

3) COMPUTATION ANALYSIS FOR DIFFERENT CONVOLUTION

In this section, we will discuss the ability to reduce computational effort between the proposed splice convolution and other convolutions theoretically

a: COMPARING WITH DEPTHWISE SEPARABLE CONVOLUTION

Figure 4 contrasts the depthwise separable convolution and splice convolution. Obviously, the splice convolution divides M channels into A and B channels, where $M = A + B$. And we set the ratio of them is 1:1. Channels are operated with 3×3 size kernel, and the reset B channels are operated with 1×1 size kernel. After batchnorm and ReLU, the separated channels are contacted together and become M size channels. Meanwhile, the depthwise separable convolution is a serial process.

Taking 1×1 to 3×3 convolution to a ratio of 1:1, in splice convolution the computational cost is given as:

$$\hat{F}_{L_K} = \frac{1}{2} \cdot D_H \cdot D_W \cdot D_K \cdot D_K \cdot D_M + \frac{1}{2} \cdot D_H \cdot D_W \cdot D_M \tag{7}$$

The total reduction in the computation as compared to depthwise separable convolution can be given as:

$$\begin{aligned} R_{SpliceConv/DwConv} &= \frac{\hat{F}_{L_K} + F_{L_1}}{F_{L_K} + F_{L_1}} \\ &= \frac{\frac{1}{2} D_K^2 + \frac{1}{2} + D_N}{D_K^2 + D_N} < 1 \end{aligned} \tag{8}$$

Obviously, our splicing convolution parameters are less than depthwise separable convolution.

b: COMPARING WITH GroupConv

Group convolution first appeared in AlexNet [23] to solve the problem of insufficient storage. The depthwise separable convolution is a special case of group convolution. In group convolution, the number of packets is defined as the input channel, then, delease convolution becomes a deep separable convolution. Obviously, the amount of computational parameters of the depthwise separable convolution will be less than or equal to the amount of the group convolution. Following Equation 8, it can be concluded that our splice convolution has a small amount of parameters than group convolution.

c: COMPARING WITH HetConv

Heterogeneous convolution (HetConv [33]) includes several heterogeneous convolution filters, which contain many convolution kernels of different sizes (for example, in a HetConv filter, 256 cores are 3×3 size,and the rest are 1×1 size). Let's define the P part, which controls the number of different types of cores in the convolution filter. For the P part, the $1 / P$ score in the total score will be used for the $K \times K$ core, and the remaining fraction ($1 - 1 / P$) will be used for the 1×1 core.

From above, the computational amount of 3×3 and 1×1 parameters of the HetConv convolution is as follows:

$$\widetilde{F}_{L_K} = (D_H \cdot D_W \cdot D_M \cdot D_N \cdot D_K \cdot D_K)/P \tag{9}$$

$$\widetilde{F}_{L_1} = (D_H \cdot D_W \cdot D_N) \cdot (D_M - \frac{D_M}{D_P}) \tag{10}$$

This shows that, when the parameter $P = D_M$, the amount of parameters of the HetConv convolution will be less than the amount of calculation of the depthwise separable convolution. However, it is worth noting that the value of D_M is the number of channels of the feature map, which is often 32, 64, 256 or even larger. According to the parameter P of the paper, the value is generally around 4, 8. If it is too large, it will affect the accuracy of the network.

When we define the ideal case, $P = D_M$, we compare the parameter calculations of the splice convolution and HetConv convolution as follows:

$$\begin{aligned} R_{SpliceConv/HetConv} &= \frac{\hat{F}_{L_K} + F_{L_1}}{\widetilde{F}_{L_K} + \widetilde{F}_{L_1}} \\ &= \frac{\frac{1}{2} D_K^2 D_M + \frac{1}{2} D_M + D_M D_N}{D_K^2 D_N + D_M D_N - D_N} < 1 \end{aligned} \tag{11}$$

It can be seen that our spliced convolution has the smallest amount of computational parameters compared to the current standard convolution, group convolution, depthwise separable convolution or HetConv convolution.

4) DISORDERED COMBINATION SPLICING

In this section, we make further optimize proposed methods via solving the boundary effect produced by each layer of splice convolution. The so-called boundary effect, that is, when we use the splice convolution (two size kernels) to do the convolution operation, if each splice position is fixed, both size of the kernel will not be flow, which will lead to a lack of channel information exchange. For example, when the ratio of 3×3 and 1×1 size kernel is 1:1, the splicing rule is that the first half is 3×3 size kernel, the latter half is 1×1 size kernel. When such a convolution operation is applied to the feature map, if the order of splice convolution is unchanged, then only the first half of the feature map of each layer is operated by the 3×3 convolution kernel, and the latter half is always only acted by the 1×1 size kernel. It is obvious that the computational complexity of 1×1 convolution is less than the convolution of 3×3, but the feature extraction effect of 3×3 size kernel is stronger than the 1×1. The consequence of this operation would be that the first half of the feature extraction would be stronger than the second half of the feature extraction that would be detrimental to the entire convolutional neural network. Therefore, in order to avoid this boundary effect, it is necessary to disrupt the sequential splicing convolution.

As shown in Figure 5, we will mess up the splice network at each layer. The first layer feature map is convoluted with a disordered splice convolution, in which the first half is 3×3 size core, and the latter half is 1×1. Then in the second layer, we reset the order of the splice convolution and place the 3×3 size kernel in the middle position. The ratio of convolution kernels of 3×3 and 1×1 is the same, that is, the number of convolution kernels of the two layers is the same for each layer, thus ensuring the same amount of parameters. By the same token, the Third layer and the next feature map all use a disordered splice convolution operation. The disordered splice convolutions have no fixed order, and we use a random way to disrupt. The first and last two layers of splicing use a fixed splicing rule. That is, the position of 3×3 and 1×1 size kernel at the input is the former in the front of the splicing, and the last layer is in the tail.

In short, the disordered splice convolution operation can effectively solve the global impact of the boundary effect. In subsequent experiments, we proved that the disordered splice convolution is more effective to enhance the performance of spliced convolutional neural networks.

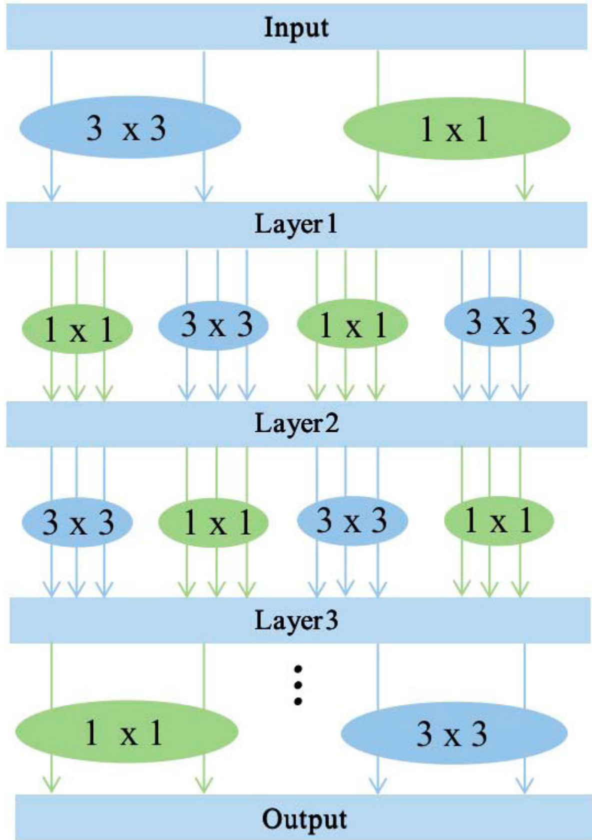


FIGURE 5. The disrupted combination spliced convolution structure of each layer is different. From top to bottom, except for the first and last two layers of convolution convolution in the order of 3×3 and 1×1 convolution, the other layers are randomly arranged in a disordered order and the ratio of the convolution kernels is the same.

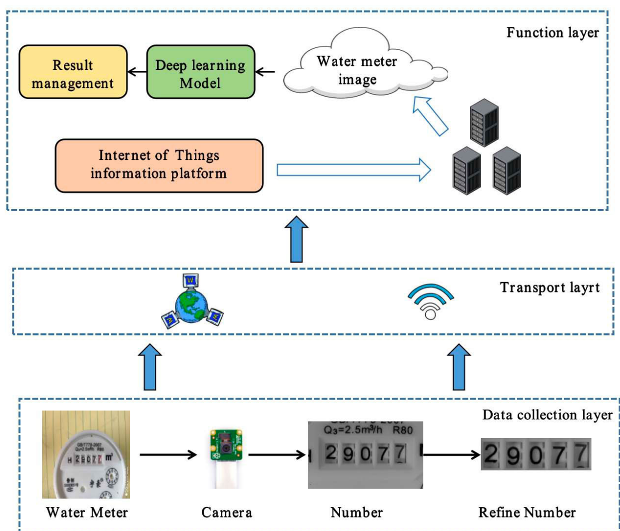


FIGURE 6. The prototype for automatic water meter reading.

IV. SMART WATER METER READING SYSTEM

We have also implemented a prototype system that can run in the real-world environment. As can be seen Figure 6, the system framework is divided into three layers from bottom to top. The bottom layer is called the data perception layer, which mainly focuses on the work of collecting data and scanning data. All the cameras are on this layer. These cameras can capture meter images and transfer these pictures

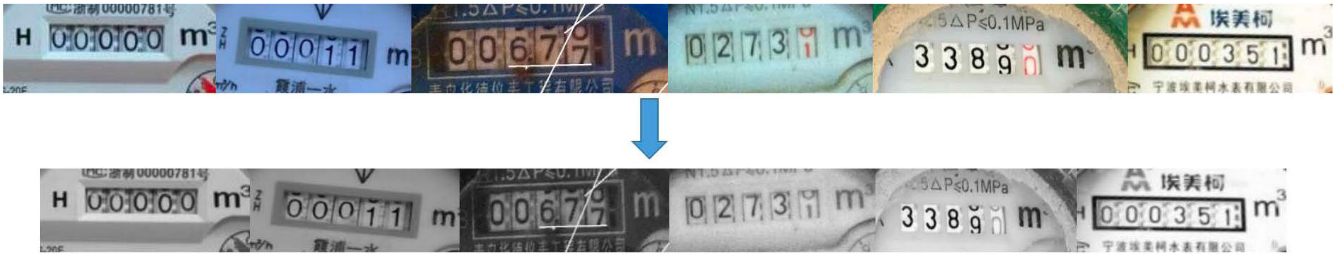


FIGURE 7. Images captured by camera on the surface of water meter and being pre-processed.

TABLE 1. The table shows the body backbone architecture of the recognition network.

operator	size	channels	repeat
Conv 3×3	224×224	32	1
SpliceConv	112×112	64	1
SpliceConv	56×56	128	2
SpliceConv	28×28	256	4
SpliceConv	14×14	512	2
SpliceConv	7×7	1024	2
Pooling & FC	7×7	–	1

to the transport layer. The transport layer is the middle layer of the system, which includes uploading the image data to the function layer through the wireless module. In practical application scenarios, we employ the Wifi or 4G-network modules for data transfer and command transfer in real-time. The top function layer is the core of the whole system. It is responsible for the real-time receiving and distributed storage of data. The proposed light-weight spliced convolution network is also run on the function layer. We have designed a convenient user interface for water companies to access and visit.

V. EXPERIMENTS

A. DATE PREPROCESSING

In the experiment, we construct the water meter dataset to verify the performance of the system. We divided the dataset into a training set and a validation set. The input of the network is captured by the camera. In order to reduce the transmission time of the network and the occupied bandwidth in our system, we will pre-process the captured pictures to the size of 160*40, and the average size of each picture is about 3.3k. Figure 7 shows an example of images, which are the original version vs. the pre-processed version. The dataset includes 6000 images.

B. NOTATIONS

In this subsection, we give some notation related to our experiment. We define our proposed spliced convolution in the following as **LW**, which means leigh-weight. When we add the numbers 4, 16, 32 after the suffix, like **LW4**, **LW16** and **LW32**, it means that we have the ratio of the 1×1 and 3×3 convolution kernels are 4:1, 6:1 and 32:1 respectively. And the absence of the number represents our original splicing convolution ratio of 1:1.

Finally, we chose the compact network body architecture shown in table 1 as the basic backbone, then use the proposed spliced convolution strategy to further reduce the network parameters and computation by replacing the traditional convolution strategy.

TABLE 2. Performance acceleration and compression for light-weight spliced convolution network.

Method	Size(M)	MFLOPs	Para Reduced(%)	Inferring time(s)
ConvNet	29.29	4997	–	0.21 ± 0.01
LW	4.21	601	85.62	0.13 ± 0.01
LW4	4.20	597	85.66	0.11 ± 0.02
LW16	4.19	594	85.69	0.10 ± 0.01
LW32	4.18	592	85.72	0.08 ± 0.005

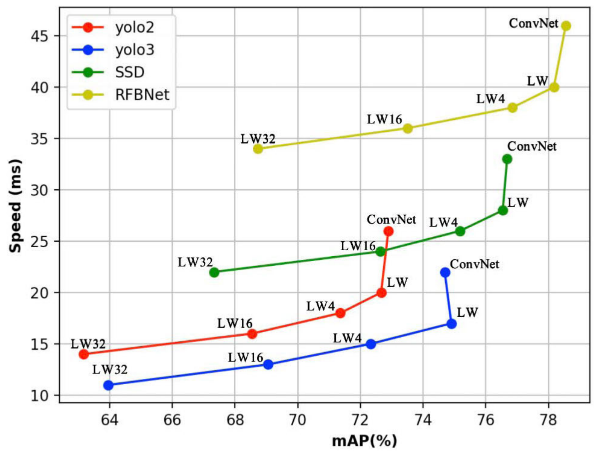


FIGURE 8. Mode speed vs. mean average precision.

C. RESULTS

1) RESULTS FOR MODEL ACCELERATION AND COMPRESSION

In this part, we estimate the running performable of the proposed model by three evaluation metrics, which are FLOPs(floating-point operations per second), model size, and inferring time, respectively. The FLOPs can be used to estimate the ability of computational consuming. The model size presents the volume of the network which is used to estimate the ability of model compression. The inferring time is running speed of model which can show the model efficiency. Table 2 shows the experimental results. By comparison, we can see that our method can reduce nearly 10× computational consuming, 7× model storage. The proposed model also saves 3× running time. It is a conclusion that the

FIGURE 9. Display of the water meter recognition system.

proposed light-weight spliced convolution network is very suitable for IoT applications.

2) RESULTS FOR WATER METER READING ACCURACY

In this part, we will estimate the system accuracy by mAP (mean Average Precision), which is a popular metric in measuring the accuracy of meter number recognition. We apply the proposed light-weight spliced convolution into four different network architectures. Figure 8 illustrates the relations between model speed and mean Average Precision. With the proportion of 1×1 convolution kernel increases, the shorter the running time, the lower the mAP. As shown in Figure 8, the proposed spliced kernel can accelerate and compress the network with light performance loss. And so we see that as well, these classic network frameworks with LW kernel achieves the nearly best performance since the disordered combination step in the proposed model has a similar function compared with dropout technology.

The Figure 9 demonstrate the running result of our reading system. The reading of the surface of the water meter is finally output as a calibration frame and corresponding digital symbols through the inspection network.

VI. CONCLUSION

In this paper, we focused on the problem of water meter recognition in smart city applications. Aiming at this task, we proposed a novel light-weight spliced convolution network to recognize the meter number, which can simplify the number of 3×3 convolutions by splicing a certain number of 1×1 and 3×3 size kernel. We then prove the superiority of our network by theoretical analysis. We conduct a real-world dataset to estimate the proposed network. The experimental results demonstrate that the proposed network has the ability to identify the water meter number accurately and simultaneously requires fewer parameters and less computation. Our model can reduce nearly $10\times$ computational consuming, $7\times$ model storage, and save $3\times$ running time. we also have implemented the prototype system which can handle real-time data base on the distributed cloud platform. In the following work, we will further optimize our algorithm and system architecture to achieve better performance from other perspectives, e.g., transplanting the algorithm to the smart camera.

ACKNOWLEDGMENT

(Chunshan Li, Yukun Su, and Rui Yuan contributed equally to this work.)

REFERENCES

[1] M. J. Mudumbe and A. M. Abu-Mahfouz, "Smart water meter system for user-centric consumption measurement," in *Proc. IEEE 13th Int. Conf. Ind. Inform. (INDIN)*, Jul. 2015, pp. 993–998.
[2] J. Wan, H. Yan, H. Suo, and F. Li, "Advances in cyber-physical systems research," *KSII Trans. Internet Inf. Syst.*, vol. 5, no. 11, pp. 1–19, Nov. 2011.
[3] N. Jazdi, "Cyber physical systems in the context of industry 4.0," in *Proc. IEEE Int. Conf. Autom., Qual. Test., Robot.*, May 2014, pp. 1–4.
[4] I. Culjak, D. Abram, T. Pribanic, H. Dzapo, and M. Cifrek, "A brief introduction to OpenCV," in *Proc. 35th Int. Conv. MIPRO*, May 2012, pp. 1725–1730.
[5] A. Lorsakul and J. Suthakorn, "Traffic sign recognition for intelligent vehicle/driver assistance system using neural network on opencv," in *Proc. 4th Int. Conf. Ubiquitous Robots Ambient Intell.*, Nov. 2007, pp. 1–6.
[6] R. Shrivastava, "A hidden Markov model based dynamic hand gesture recognition system using OpenCV," in *Proc. 3rd IEEE Int. Advance Comput. Conf. (IACC)*, Feb. 2013, pp. 947–950.
[7] X.-D. Zhou, D.-H. Wang, F. Tian, C.-L. Liu, and M. Nakagawa, "Handwritten Chinese/Japanese text recognition using semi-Markov conditional random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2413–2426, Oct. 2013.
[8] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven, "Photoocr: Reading text in uncontrolled conditions," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 785–792.
[9] P. He, W. Huang, Y. Qiao, C. C. Loy, and X. Tang, "Reading scene text in deep convolutional sequences," in *Proc. 30th AAAI Conf. Artif. Intell.*, May 2016, pp. 3501–3508.
[10] F. Yang, L. Jin, S. Lai, X. Gao, and Z. Li, "Fully convolutional sequence recognition network for water meter number reading," *IEEE Access*, vol. 7, pp. 11679–11687, 2019.

[11] M. Suresh, U. Muthukumar, and J. Chandapillai, "A novel smart water-meter based on IoT and smartphone app for city distribution management," in *Proc. IEEE Region Symp. (TENSYMP)*, Jul. 2017, pp. 1–5.

[12] Y. Yin, L. Chen, Y. Xu, J. Wan, H. Zhang, and Z. Mai, "Qos prediction for service recommendation with deep feature learning in edge computing environment," *Mobile Netw. Appl.*, pp. 1–11, 2019.

[13] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security—A survey," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1802–1831, Dec. 2017.

[14] Y. Yin, Y. Xu, W. Xu, M. Gao, L. Yu, and Y. Pei, "Collaborative service selection via ensemble learning in mixed mobile network environments," *Entropy*, vol. 19, no. 7, p. 358, 2017.

[15] C. S. Gastouniotis, N. Bandeira, and K. C. Wilson, "Automated remote water meter readout system," U.S. Patent 4 940 976, A, Jul. 10, 1990.

[16] Y.-W. Lee, S. Eun, and S.-H. Oh, "Wireless digital water meter with low power consumption for automatic meter reading," in *Proc. Int. Conf. Converg. Hybrid Inf. Technol.*, Aug. 2008, pp. 639–645.

[17] C. Rong-Su and L. Zhi-Ming, "A novel method of word wheel's mumeral recognition and automatic meter reading system," *Elect. Meas. Instrum.*, vol. 8, pp. 28–36, Aug. 2007.

[18] G. Ju and Y. Hua, "An effective two-times recognition algorithm of meter digital image," *J. Southeast Univ., Natural Sci. Ed.*, vol. 43, no. SI, pp. 153–157, 2013.

[19] X. R.-P. Z. Rong-Chun and Y. S.-C. Y. Fei, "An automatic reading and recognition system for water meter based on DSP and image processing," *Meas. Control Technol.*, no. 5, pp. 21–25, 2007.

[20] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "ICNet for real-time semantic segmentation on high-resolution images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 405–420.

[21] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 325–341.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21–37.

[26] S. Liu and D. Huang, "Receptive field block net for accurate and fast object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 385–400.

[27] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proc. Int. Conf. Pattern Recognit.*, vol. 3, Aug. 2006, pp. 850–855.

[28] R. Smith, "An overview of the tesseract OCR engine," in *Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 2. Sep. 2007, pp. 629–633.

[29] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: https://arxiv.org/abs/1312.4400

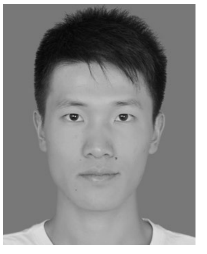
[30] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.

[31] S. Deng, Z. Xiang, J. Yin, J. Taheri, and A. Y. Zomaya, "Composition-driven IoT service provisioning in distributed edges," *IEEE Access*, vol. 6, pp. 54258–54269, 2018.

[32] Y. Yin, W. Xu, Y. Xu, H. Li, and L. Yu, "Collaborative QoS prediction for mobile service with data filtering and SlopeOne model," *Mobile Inf.Systems*, vol. 2017, Jun. 2017, Art. no. 7356213.

[33] P. Singh, V. K. Verma, P. Rai, and V. P. Namboodiri, "HetConv: Heterogeneous Kernel-based convolutions for deep CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4835–4844.

CHUNSHAN LI is currently an Assistant Professor with the Department of Computer Science, Harbin Institute of Technology, Weihai, China. His research interests include artificial intelligence and data mining.



YUKUN SU received the B.S. degree in software engineering from the South China University of Technology, in 2017, where he is currently pursuing the master's degree with the School of Software Engineering. His research interests include computer vision and AIOT.

RUI YUAN received the B.S. degree from the School of Communication and Design, Sun Yat-sen University, in 2011, and the M.S. degree from the School of Journalism and Communication, The Chinese University of Hong Kong, in 2013. His research interests include machine learning and data mining.

DIANHUI CHU received the Ph.D. degree from the Department of Computer Science and Technology, Harbin Institute of Technology, Weihai. He is currently a Professor with the Harbin Institute of Technology. His research interests include within the domain of service computing, service engineering, and software architecture.

JINHUI ZHU is currently an Associate Professor with the School of Software Engineering, South China University of Technology, China. His research interests include artificial intelligence and robotics.