A Real-Time Forest Fire Recognition Method Based on R-shufflenetv2

Mengna Li[1], *Youmin Zhang*[2*], *Lingxia Mu*[1], *Jing Xin*[1], *Xianghong Xue*[1], *Shangbin Jiao*[1], *Han Liu*[1], *Guo Xie*[1], *and Yingmin Yi*[1]

[1]Shaanxi Key Laboratory of Complex System Control and Intelligent Information Processing
Xi'an University of Technology, Xi'an, Shaanxi 710048, China
[2]Department of Mechanical, Industrial and Aerospace Engineering, Concordia University, Montreal, Quebec H3G 1M8, Canada
*Email: youmin.zhang@concordia.ca

2022 5th International Symposium on Autonomous Systems (ISAS) | 978-1-6654-8708-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISAS55863.2022.9757053

Abstract—**The recognition of early forest fires can reduce the resource loss caused by fire combustion. A real-time forest fire image recognition method based on r-shufflenetv2 network is proposed. R-shufflenetv2 is mainly composed of a series of r-shufflenetv2 units and is an improved version of the shufflenetv2 network. In order to improve the recognition accuracy, the shufflenetv2 unit is reconstructed by using the residual structure, which enhances the feature extraction ability of the network. The experimental results show that the r-shufflenetv2 network is a good fire recognition model. On the benchmark FLAME dataset, r-shufflenetv2 has higher fire recognition accuracy than the original shufflenetv2. In addition, r-shufflenetv2 also achieves real-time detection, and its recognition speed is about 31 FPS.**

Keywords—**Forest Fire, Fire Recognition, Real-Time Detection, Shufflenetv2, Residual structure.**

I. Introduction

Ecological crises caused by forest fires are usually fatal. On the one hand, forest burning will destroy vegetation, pollute the environment, endanger the survival of wild animals and plants and the development of human society [1]. On the other hand, large-scale forest fires are often difficult to rescue manually, and can only wait to be extinguished. However, if forest fires can be detected at an early stage, it can not only help firefighters to put out the fires, but also reduce the huge loss of resources caused by fire burning.

With the development of sensor technology, it is popular to monitor and forecast forest fire based on optical sensor, especially using RGB camera to monitor fire [2]. Therefore, researchers have developed many fire recognition methods based on RGB images. At present, fire recognition methods can be divided into two categories: *traditional fire image recognition methods* and *deep learning-based fire image recognition methods*, according to whether it is necessary to manually extract fire feature information [3].

Traditional fire image recognition methods use hand-designed fire features to identify forest fires. These features usually consist of the color, texture, frequency and motion features of forest flame or smoke, etc. [4]. Generally speaking, the implementation process of traditional methods mainly includes four parts: *image preprocessing, extraction of suspected fire areas, fire feature extraction and fire recognition and classification* [5]. Image preprocessing enhances the target region by removing the interference of noise in the image to improve the subsequent recognition performance. Extraction of suspected fire areas can separate the suspected fire area from the background or other objects, so as to avoid the whole image operation and reduce the calculation amount. Fire feature extraction describes the phenomenon of forest fires by extracting effective fire feature information. Fire identification and classification is to identify and classify the extracted fire features by using classification models such as support vector machine (SVM) and artificial neural network (ANN).

The main advantages of the above traditional methods are that the amount of calculation is small, the recognition speed is fast, and the task of fire recognition in a simple environment (ie, the background is simple and the occlusion is small) can be realized. However, in forests, due to factors such as obstacles and light changes, early fires are difficult to be captured by video surveillance and identified, making it difficult for traditional methods to be applied to forest fire identification in this environment. On the other hand, the artificially designed fire features are blind and complex, and sometimes cannot accurately describe the forest fire phenomenon, resulting in a low fire recognition rate of traditional methods.

Deep learning-based fire image recognition methods can automatically extract fire features from images using convolutional neural networks (CNN), avoiding the use of hand-designed fire features. Compared with traditional methods, these fire features extracted using CNNs are more efficient, so it can more accurately describe forest fire phenomena and improve the accuracy of fire image recognition [6]. In [7], in order to balance the efficiency and accuracy of the model, the authors take advantage of the lightweight network GoogleNet to propose a CNN architecture for fire detection in surveillance videos. In [8], the author replaces the backbone feature network of YOLOv4 with MobileNet, thereby proposing a lightweight network structure, YOLOv4-Light. Experiments show that this method has higher FPS and fewer network parameters in forest fire detection. In [9], the authors adopted a lightweight network design and channel pruning method to achieve accurate and fast detection of smoke and flame images.

From the above, in forest fire detection, more use of lightweight convolutional neural networks can not only improve the processing speed of the algorithm, but also reduce network parameters so that the network can be deployed on some edge processors.

Shufflenetv2 is a lightweight network with good comprehensive performance [10], which has reached a relatively good level in terms of speed and accuracy, and is widely used in many different fields to solve practical problems [11, 12, 13].Therefore, this paper studies the Shufflenetv2 network and proposes to use it to identify forest fire images. In addition, in order to improve the fire recognition accuracy, residual learning is introduced to modify the structure of the shufflenetv2 network, thereby proposing the r-shufflenetv2 network. Experimental results show that the r-shufflenetv2 network has better recognition accuracy than the original Shufflenetv2 on the benchmark FLAME (Fire Luminosity Airborne-based Machine Learning Evaluation) dataset provided by [14]. At the same time, the r-shufflenetv2 network realizes real-time fire identification, and its running speed reaches about 31FPS.

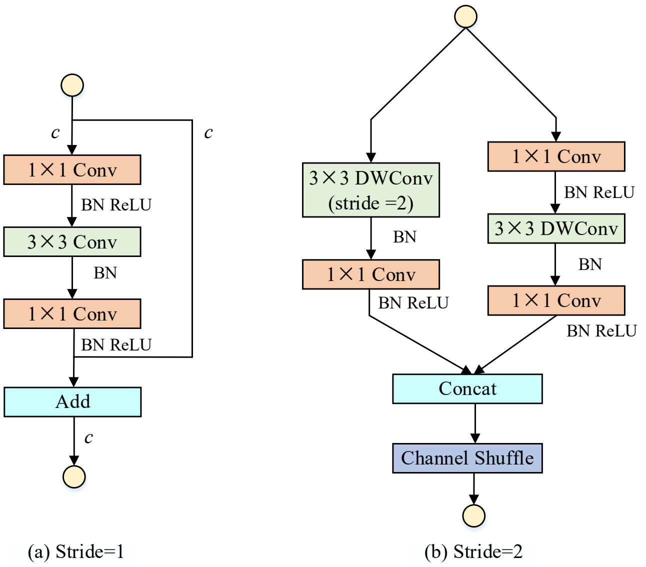
The rest of this paper is organized as follows. The second section introduces a real-time forest fire image recognition method based on r-shufflenetv2 network. Section III presents some fire recognition results of this method, and analyzes the performance and efficiency of the network. Section IV summarizes the full-text work and possible future research work.

II. RCOGNITION METHOD

This section introduces a real-time fire identification method based on the r-shufflenetv2 network. First, how to design the required r-shufflenetv2 unit is described. Then, the r-shufflenetv2 network is built using the designed r-shufflenetv2 unit, and the overall architecture of the network is shown.

A. R-shufflenetv2 unit

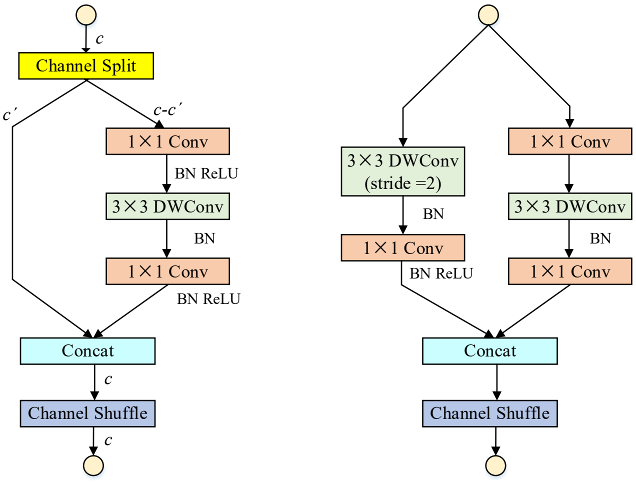
Fig. 1 shows the composition of the r-shufflenetv2 unit. According to the convolution stride of 1 or 2, it is divided into two different structures. When the stride is 2, it indicates that spatial downsampling of the input feature map is required. At this time, the size of the feature map is reduced to one-half of the original size.



(a) Stride=1 (b) Stride=2

Fig. 1. The r-shufflenetv2 unit.

The biggest difference between the r-shufflenetv2 unit and the original shufflenetv2 unit is that the structure of the unit with a convolution stride of 1 is different. In this paper, the residual structure is used to reconstruct the r-shufflenetv2 unit with a convolution stride of 1. Because the residual structure can learn the difference between input features and output features, solve the problem of network degradation, and make the network have better feature expression ability. This has a positive effect on the improvement of network recognition accuracy. To facilitate the comparison of the two, this paper shows the structure of the shufflenetv2 unit, as shown in Fig. 1.



(a) Stride=1 (b) Stride=2

Fig. 2. The shufflenetv2 unit.

Stride=2 Unit. According to Fig. 1(b) and Fig. 2(b), the structure of the unit with the convolution stride of 2 in r-shufflenetv2 is the same as that of the original shufflenetv2 unit, so it follows the design of the efficient network architecture proposed in [10] the four principles. The four principles are as follows (1) Keeping the same number of input and output feature channels can minimize memory access cost; (2) Excessive use of group convolution will increase memory access cost; (3) Excessive fragmentation of the network will reduce its parallelism; (4) Element-wise operations will increase memory access costs. Since the unit with convolution stride of 2 in r-shufflenetv2 is proposed based on the above principles, it has a small memory access cost, so the intuitive result is that the running speed of the network will increase.

Stride=1 Unit. According to Fig. 1(a), the unit with convolution stride of 1 in r-shufflenetv2 is composed of residual structure. The reasons for using the residual structure are as follows. In the original shufflenetv2 unit (shown in Fig. 2(a)), at the beginning of each unit, the feature channel of the input feature is split into two branches, namely $c-c'$ channels and c' channels, and $c'=c/2$ in this paper. For the convenience of explanation, it is assumed that branch c' maintains the identity mapping, that is, does not do any convolution processing. Therefore, after going through this branch, the number of input features and output features remains the same. Branch $c-c'$ uses convolution operations, including standard convolution and depthwise convolution (DWConv as shown in Fig. 2), to extract image features. Moreover, after going through branch $c-c'$, the number of input features and output features is still the same. After through channel split, the convolution operation on the above two branches is similar to the residual structure. Therefore, this paper reconstructs the unit with a convolution stride of 1 using the residual structure. Finally, the element-wise addition operation ensures that the input features and output features have equal feature channels.

Due to the use of a residual structure, the structure of a unit with a convolution stride of 1 violates the four principles mentioned above. This increase network parameters and memory access cost, which in turn affects the running speed of the network. However, residual structure can learn the difference between input and output and its unique skip structure [15], so it has the advantages of easy optimization, mitigation of gradient disappearance, protection of information integrity, etc. Therefore, using the residual structure can enhance network performance, thereby improving the recognition accuracy of the model for fire images. This is also a trade-off between speed and accuracy.

B. Network architecture

This paper uses the above r-shufflenetv2 unit to build the r-shufflenetv2 network, and its overall architecture follows the original shufflenetv2 architecture.

As shown in Table I, the first layer of the r-shufflenetv2 network is composed of 3×3 standard convolutions with 24 convolutional filters. The second layer is the maxpooling layer, which reduces the feature size by downsampling. It helps to simplify the network complexity and reduce the amount of computation. The next few layers are composed of a series of r-shufflenetv2 units. These r-shufflenetv2 units are divided into three distinct stages according to the size of the input features. The convolution stride of the first r-shufflenetv2 unit in each stage is 2, and the convolution stride of the remaining units is 1. Within each stage, keep other hyperparameters unchanged. The last three layers consist of convolutional layers, global pooling layers, and fully connected (FC) layers. The purpose of convolution and global average pooling is to downsample the output features of the final stage into feature vectors of feature size $1\times1\times1024$. The purpose of the FC layer is to map this feature vector to two different outputs (fire and no fire).

TABLE I. Architecture of r-shufflenetv2 network

Layer	Output size	KSize	Stride	Repeat	Output channels
Image	224×224	-	-	-	3
Conv1	112×112	3×3	2	1	24
MaxPool	56×56	3×3	2	1	24
Stage2	28×28	-	2	1	116
	28×28	-	1	3	
Stage3	14×14	-	2	1	232
	14×14	-	1	7	
Stage4	7×7	-	2	1	464
	7×7	-	1	3	
Conv5	7×7	1×1	1	1	1024
GlobalPool	1×1	7×7	-	-	1024
FC	-	-	-	-	2

Note that when stride=2, the sliding distance of each step of the convolution kernel on the feature map is 2. At this time, the size of the feature map is reduced to one-half of the original size. In addition, in CNN, the stride of the convolution generally does not exceed 3. This is because as the stride increases, the fewer features will be extracted; on the contrary, the more features will be extracted.

III. Experiment

This section mainly uses quantitative analysis and qualitative analysis to evaluate the recognition performance of the proposed network. In the quantitative analysis, three evaluation indexes are introduced, and the recognition accuracy and running speed of the proposed network are measured according to these evaluation indicators. In the qualitative analysis, the proposed network is used for forest fire image recognition, and the network performance is analyzed according to the recognition results.

A. Qualitative analysis

1) Evaluation indexes

Accuracy: Accuracy represents the proportion of the number of samples correctly predicted to the total number of samples, calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

where TP is the number of true positive samples, representing the samples that belong to positive category and are correctly predicted. TN is the number of true negative samples, representing the samples that belong to negative category and are correctly predicted. FP is the number of false positive samples, which represents the samples that belong to negative category but are wrongly predicted as positive category. FN is the number of false negative samples, which represents the samples that belong to positive class but are wrongly predicted as negative class.

The higher the value of accuracy, the better the model performance.

F1 score: F1 score is an index to measure the correctness of positive prediction, which represents the proportion of positive samples among the samples marked as positive. F1 score is regarded as a harmonic average of the model precision and recall, which ranges from 0 to 1. The calculation of F1 score is as follows:

$$F1 = 2\times\frac{P\times R}{P+R} \tag{2}$$

$$P = \frac{TP}{TP+FP} \tag{3}$$

$$R = \frac{TP}{TP+FN} \tag{4}$$

where, P represents precision and R represents recall.

The higher the value of F1 score, the better the model performance.

FPS: FPS is the number of frames processed per second, calculated as follows:

$$FPS = \frac{frame}{time} \tag{5}$$

It is generally believed that real-time recognition is achieved when the network is running above 20 FPS.

In this paper, accuracy and F1 score are used to measure the recognition accuracy of the network, and FPS is used to measure the processing speed of the network.

2) Results analysis

This paper uses the FLAME dataset to train and test the proposed r-shufflenetv2 network. The FLAME dataset is an aerial forest fire dataset developed in [14]. The dataset contains 47,992 images, divided into train set and test set. Images on the train set were acquired using a DJI Matrice 200 UAV and Zenmouse X4S camera, and images on the test dataset were acquired using a DJI Phantom 3 UAV and its default camera. There is no overlap between training samples and test samples. This confirms that the method network in this paper is not biased towards the characteristics of the imaging device.

In the experiments, data augmentation methods (including rotation, cropping, etc.) are used to address the imbalanced number of samples. The number of train is 40 epochs, and batch size is 32. The Adam optimizer is used during training, and the learning rate is set to 0.001 and kept constant. All networks are trained and tested with Windows 10 system using i7-9700k and Nvidia RTX 2080 Ti. Table II reports the accuracy, F1 score and running speed of shufflenetv2 network and r-shufflenetv2 network on the FLAME dataset.

TABLE II. Accuracy, F1 score and running speed of shufflenetv2 and r-shufflenetv2 on the benchmark FLAME dataset.

Performance	Accuracy (%)		F1 Score	Running speed
	Train set	Test set		
shufflenetv2	99.81	82.12	0.8544	34 FPS
r-shufflenetv2	99.63	86.33	0.8908	31 FPS

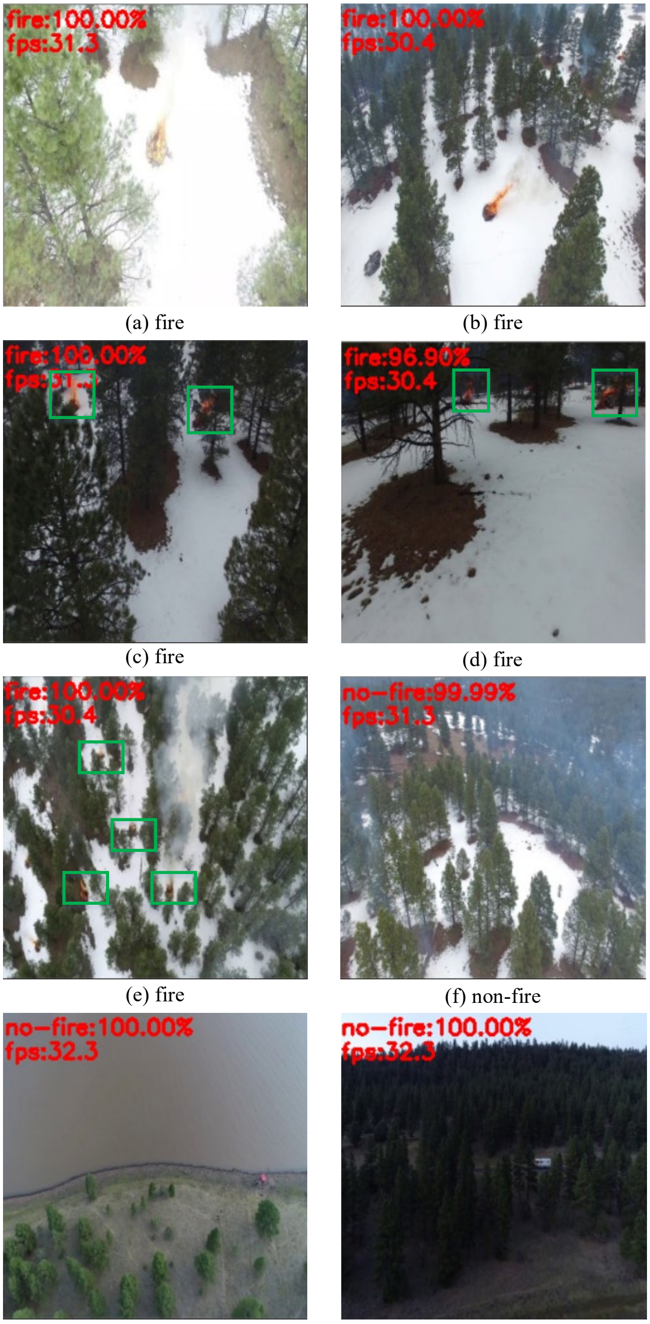
According to Table II, on the benchmark FLAME train set, the recognition accuracy and F1 score of the r-shufflenetv2 network are slightly lower than those of the shufflenetv2 network. This shows that for the same imaging device, the recognition performance of the original shufflenetv2 network is slightly better. However, on the benchmark FLAME test set, the accuracy of the r-shufflenetv2 network is 4.21% higher than that of the shufflenetv2 network. This indicates that the proposed network has better generalization ability and its recognition effect is not biased towards the features of any imaging device. At the same time, it also shows that the robustness of the proposed network is better than that of the original shufflenetv2 network, so it is suitable for fire image recognition on different imaging devices. In addition, on the benchmark test set, the F1 score of the r-shufflenetv2 network is higher than that of the shufflenetv2 network, which shows that r-shufflenetv2 can better balance the precision and recall of the model.

Therefore, it can be concluded that the recognition performance of r-shufflenetv2 network is better than shufflenetv2 on the benchmark FLAME dataset. The main reason is the use of residual structure, which increases the feature expression ability of the network.

As can be seen from Table II, the running speed of the r-shufflenetv2 network is about 31 FPS. Although it is lower than the running speed of shufflenetv2 network, it also realizes real-time recognition. The reason why the running speed of the r-shufflenetv2 network is lower than that of the shufflenetv2 network is that the residual structure is used in the network construction, which increases the network parameters and memory access costs, and violates the design principles proposed in [10].

B. *Qualitative analysis*

Fig. 3 shows the results of identifying aerial forest fire images using the r-shufflenetv2 network, in which the green boxes indicate some fire areas obscured by trees.



(a) fire (b) fire

(c) fire (d) fire

(e) fire (f) non-fire

(g) non-fire (h) non-fire

Fig. 3. Recognition results of forest fire images.

As can be seen from Fig. 3, the proposed network can identify some early forest flames and smoke, with the ability to distinguish fire images from non-fire images. Secondly, this method is sensitive to the recognition of fire areas covered by trees (as shown in Fig. 3(c)-3(e)). Finally, the recognition results in Fig. 3(f) show that r-shufflenetv2 has the ability to distinguish forest smoke and fog.

Quantitative and qualitative results show that the r-shufflenetv2 network is a well-performing fire recognition network that can be used to identify early forest fire images. Moreover, the r-shufflenetv2 network can realize real-time fire detection, and its processing speed reaches about 31FPS.

IV. Conclusion and Future Work

This paper proposes a real-time forest fire image recognition method based on r-shufflenetv2 network. The r-shufflenetv2 network is mainly composed of a series of r-shufflenetv2 units and is an improved version of the shufflenetv2 network. The r-shufflenetv2 unit contains two different structures, divided according to whether the convolution stride is 1 or 2. In order to improve the recognition accuracy, this paper uses a residual network to reconstruct the r-shufflenetv2 unit with a convolution stride of 1, thus enhancing the feature expression ability of the network. The experimental results show that the r-shufflenetv2 network is an excellent fire recognition model, which improves the accuracy of fire recognition on the benchmark FLAME dataset. The r-shufflenetv2 network achieves real-time fire detection with a recognition speed of about 31 FPS.

In the future, we intend to deploy the proposed network in edge computing devices, such as UAV onboard counter processors, to realize real-time UAV-based fire image detection. On the other hand, we intend to process and fuse the fire information obtained by multiple sensors to improve the efficiency of forest fire detection.

Acknowledgment

This work is partially supported by the National Natural Science Foundation of China (Grants 61833013, 61873200, 61903297, and 62003162), Natural Science Foundation of Shaanxi Provincial Department of Education (19JK0569), and Natural Sciences and Engineering Research Council of Canada.

References

[1] V. Chowdary, M. K. Gupta, and R. Singh, “A review on forest fire detection techniques: A decadal perspective,” *International Journal of Engineering and Technology*, 2018. 7(3.12): 1312-1316.

[2] Alkhatib, and A. A. Ahmad. “A review on forest fire detection techniques.” *International Journal of Distributed Sensor Networks*, 2014.

[3] X. Xia, F. N. Yuan, L. Zhang, L. Z. Yang and J. T. Shi, “From traditional methods to deep ones: Review of visual smoke recognition, detection, and segmentation,” *Journal of Image and Graphicss*, 2019, 24(10): 1627-1647.

[4] C. Yuan, Z.X. Liu and Y.M. Zhang, “Vision-based forest fire detection in aerial images for firefighting using UAVs,” *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2016, pp. 1200-1205.

[5] J. T. Shi, F. N. Yuan and X. Xia, “Video smoke detection: A literature survey,” *Journal of Image and Graphics*, 2018, 23(3): 303-322.

[6] S. Geetha, C. S. Abhishek and C. S. Akshayanat, “Machine vision based fire detection techniques: A survey,” *Fire Technology*, 2021, 57: 591-623.

[7] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho and S. W. Baik, “Convolutional neural networks based fire detection in surveillance videos,” in *IEEE Access*, 2018, 6: 18174-18183.

[8] R. Fan and M. Pei, “Lightweight forest fire detection based on deep learning,” *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*, 2021, pp. 1-6.

[9] L. Zeng, G. Xu, L. Dong and P. Hu, “Fast smoke and flame detection based on lightweight deep neural network,” *2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2020, pp. 232-235.

[10] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, “ShuffleNet V2: Practical guidelines for efficient CNN architecture design,” *2018 IEEE European Conference on Computer Vision (ECCV)*, 2018.

[11] R. Yang, X.Y. Lu, J. Huang, J. Zhou, J. Jiao, and Y.F. Liu, et al. , “A multi-source data fusion decision-making method for disease and pest detection of grape foliage based on ShuffleNetV2, ” *Remote* Sensing, 2021(13):5102.

[12] H. Ran, S. Wen, S. Wang, Y. Cao, P. Zhou and T. Huang, “Memristor-based edge computing of ShuffleNetV2 for image classification, ” in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2021, 40(8): 1701-1710.

[13] X. Li, J. Li and Z. Lv, “ShuffleNet2MC: A method of light weight fault diagnosis, ” *2021 International Conference on Computer Engineering and Application (ICCEA),* 2021, pp. 264-270.

[14] A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P. Z. Fulé, and E. Blasch, “Aerial imagery pile burn detection using deep learning: The FLAME dataset,” *Computer Networks*, 2021, 193: 108001.

[15] K. M. He, X. Y. Zhang, S. Q. Ren and J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778.