ATT Squeeze U-Net: A Lightweight Network for Forest Fire Detection and Recognition

**JIANMEI ZHANG, HONGQING ZHU, (Member, IEEE), PENGYU WANG,
AND XIAOFENG LING, (Member, IEEE)**
School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China
Corresponding author: Hongqing Zhu (hqzhu@ecust.edu.cn)

ABSTRACT Forest fire is becoming one of the most significant natural disasters at the expense of ecology and economy. In this article, we develop an effective SqueezeNet based asymmetric encoder-decoder U-shape architecture, Attention U-Net and SqueezeNet (ATT Squeeze U-Net), mainly functions as an extractor and a discriminator of forest fire. This model takes attention mechanism to highlight useful features and suppress irrelevant contents by embedding Attention Gate (AG) units in the skip connection of U-shape structure. In this way, salient features are emphasized so that the proposed method could be competent at forest fire segmentation tasks with a small number of parameters. Specifically, we first replace classical convolution layer by a depthwise one and engage a Channel Shuffle operation as a feature communicator in the Fire module of classical SqueezeNet. Then, this modified SqueezeNet is employed as a substitution of the encoder of Attention U-Net and a corresponding DeFire module designed is combined into the decoder as well. Finally, to classify true fire, we take use of a fragment of the encoder in ATT Squeeze U-Net. The experimental results of modified SqueezeNet integrated Attention U-Net show that a competitive accuracy at 0.93 and an average prediction time at 0.89 second per image are achieved for reliable real-time forest fire detection.

I. INTRODUCTION

Early detection and identification of forest fire can avoid damaging disaster. Fire detection methods such as satellite-based detection, optical sensing, wireless sensing and remote sensing gain notable improvements to forest fire alarm. In this study, we focus on monitoring fire detection driven by computer vision. Computer vision mechanisms for fire detection could be mainly classified into two categories, traditional image processing method and deep Convolutional Neural Network (CNN) method. Existing conventional detection algorithms mainly operate based on visual properties of fire, such as color, spectral, texture, motion and geometric features. Despite the low cost and simplicity, traditional methods strongly rely on appropriate feature description of fire. Some natural phenomena, such as sunset and fog would cause false alarm and missing report to these approaches occasionally. To solve these problems, a more advanced fire detection scheme proposing the use of CNN technology

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

instead of feature description has attracted more and more attention. Meanwhile, recent development of GPU allows the use of CNN-based methods for flame detection. Common disadvantage of them seems to be that large datasets are required to learn the best features. As a result of this, the model would over-fitted under huge training dataset, whereas applying a small number of dataset for learning would be insufficient. Recently, some lightweight compression networks [1] that could achieve real-time processing have been introduced when reasonable mistakes are allowed.

In this article, we propose an efficient neural network architecture for forest fire detection and recognition based on Attention U-Net and SqueezeNet (ATT Squeeze U-Net). The proposed framework consists of two stages, a segmentation module extracting the shape of forest fire, and a classification module identifying whether the detected fire area is true or not. We remove the encoder of conventional Attention U-Net [2] and place our modified SqueezeNet with new designed Fire modules at the contraction path. Besides, some implementations are carried out on traditional Fire module [3] in SqueezeNet. We replace the original 3×3 convolution

layer with a Depthwise Convolution (DWConv) kernel, and a Channel Shuffle operation is added to have feature communication enhanced. We design corresponding DeFire modules for ATT Squeeze U-Net model and embed them into the decoder for better up-sampling. We then develop a new classification framework for fire identification by reusing a part of the encoder of ATT Squeeze U-Net. A discussion of how many output feature maps of the encoder layers are chosen is raised to reach a most effective selection for subsequent fire recognition. We evaluated the ATT Squeeze U-Net on some publicly available datasets, and the results demonstrate that the proposed framework can produce better fire area extraction results than some existing algorithms.

The main contributions of this study are summarized as follows:

- Fire segmentation and recognition could both be achieved at one time by the proposed ATT Squeeze U-Net.
- The SqueezeNet fragment firstly substituted in the encoder of Attention U-Net by this article significantly decrease model parameters.
- We first set DWConv and Channel Shuffle operation in Fire modules and DeFire modules of SqueezeNet, and thus improve feature learning ability while suppressing computation.
- The architecture of this proposed network may benefit other segmentation and recognition studies as well as more complex fire detection tasks.

The rest of this article is arranged as follows: Section 2 introduces related works on forest fire detection in recent years. Section 3 describes the datasets used. In Section 4, we demonstrate the proposed network architecture in detail. We discuss and analyze some experimental results in Section 5. In the last section, the main conclusions of this study and future research direction are raised.

II. RELATED WORKS

Literatures indicate that forest fire detection techniques have mainly three branches nowadays, sensor-based methods, feature-based extraction methods and deep CNN-based schemes. Previous studies are more likely to rely on the strategy of features, such as fire specific chromatograms, shape and textures, fire motion, etc. A major issue of these methods is the complex manual feature extraction tasks. Hence, recent researches extensively develop the use of deep CNN on early flame detection and have shown increasing accuracy with greatly minimized false alarm rates.

A. SENSOR BASED FIRE DETECTION

Current sensor-based fire detection designs have been introduced to early stages of the detection and provide suppression for monitoring system. These fire alarm sensors mainly include temperature sensors [4], smoke sensors [5], infrared sensors [6], optical sensors [7], gas sensors [8], etc. Qiu *et al.* [5] proposed a fire detection system using laser spectroscopic carbon monoxide sensor. They adopted a highly effective micro-controller and simple digital lock-in amplifier (DLIA)

for early fire warning. Li *et al.* [4] raised a early fire detection study called long-range Raman distributed fiber temperature sensor (RDFTS), which achieved maximum sensing distance at 30km and a spatial resolution of 28m. Although temperature sensor could provide single-point measurement, it is not available for long-distance sensing. A photoacoustic gas sensor adopting a near-infrared tunable fiber laser based on wavelength modulation spectroscopy technique is reported by Wang and Wang [8]. This sensor provides rapid and concentrated measurements of combustion products, especially C2H2, CO, and CO2 under atmospheric pressure. Besides, chemical gas sensors tend to respond quicker than smoke particle detectors. However, sensor-based fire detection system is impractical due to the requirement of regular distributed sensors in close proximity.

B. FEATURE BASED FIRE DETECTION

The forest fire detection algorithms by means of color feature are widely reported in literatures. Marbach *et al.* [9] investigated YUV color space and motion features for fire detection. This method cannot obtain real-time detection due to numerous parameters and high computational complexity. Foggia *et al.* [10] combined color, shape and motion features for real-time fire detection. Generally, color-driven methods are not effective enough since the equivalent sensitive to cloud and brightness as well, and are prone to confuse moving targets similar to flame color with real flame.

Texture description operators, especially local binary pattern (LBP), are often used to analyze texture images with flame. The success of these methods strongly relies on the identification of effective forest fire texture features. For example, Yuan [11] connected histogram sequence of LBP with local binary pattern variance pyramid and extracted fire texture feature for flame detection. Another dynamic texture descriptor with hidden Markov tree and surfacelet transform was presented by Ye *et al.* [12] as well.

Apart from the literatures considering color and texture features mentioned above, some other previous wide-fire detection algorithms were built based on features of fire shape and fire-color moving objects. A recent algorithm considered the use of shape invariant features [13]. Generally, shape variant features used in the methods may result in the reduction of generalization performances. Motion of flame exist in forest fire, hence detection algorithms based on motion behavior of fire have been reported. For instance, in [14], Yu *et al.* addressed a video fire detection algorithm using color and motion features. Mueller *et al.* [15] proposed a computational vision-based flame detection model via exploring motion features based on motion estimators. Different from rigid objects with clear contour, forest fire has diverse shape, color and moving direction varying throughout time. As a result, it is difficult for those detection algorithms to build based on the features of extracted from fire.

C. DEEP CNN BASED FIRE DETECTION

In recent years, deep CNN has been successfully applied to forest fire detecting and identifying, and has demon-

strated superior performance on detection tasks. Yin *et al.* [16] used Recurrent Neural Network (RNN) architecture to capture smoke area and motion context information. Muhammad *et al.* [17] introduced a fire detection CNN model for surveillance videos. Frizzi *et al.* [18] investigated a nine-layer CNN structure for fire or smoke detection. A number of variations of CNN algorithms have been proposed for fire detection tasks, including Region-based Convolutional Neural Network (R-CNN) method [19] and Faster R-CNN [20], etc. With the rise of target detection approaches, fire detection is no longer satisfied with determining fire, but rather locate and extract exact areas. For example, a Faster R-CNN network proposed by Barmpoutis *et al.* [21] was used to locate candidate regions for fire detection. In [22], Jiao *et al.* reported a forest fire detection model by applying YOLOv3 to UAV-based aerial images. Recently, Dunnings and Breckon [23] introduced the automatic detection system of fire region in video imagery using AlexNet [24] and superpixels. Because AlexNet uses five convolution kernels and adds three fully connected layers at the end of the network, the network contains a total of 61 million parameters. Matlani and Shrivastava [25] presented a deep feature synthesis method based on VGG-Net for the classification of smoke. VGG-Net family [26] is a series of network architectures which different depth of layers are engaged respectively. For example, VGG-19 has 19 layers and 138 million parameters, which is even larger than AlexNet. Increasing number of convolution layers is becoming a general trend of neural network design, since detection accuracy could be improved to some extent. However, high computational and large memorial cost greatly hinder deep networks from applications. More recently, some light-weight CNN architectures have been reported in fire detection systems. SqueezeNet [3] is a very successful example for wild-fire and smoke detection. Peng and Wang *et al.* [27] presented a novel CNN liked compression model SqueezeNet that reduces parameters by replacing convolutional layer with Fire module, while classification accuracy remains similar to AlexNet. Apart from SqueezeNet, many other studies on light-weight networks have raise concerns. Muhammad *et al.* [28] introduced an effective MobileNet-based early fire detection framework for surveillance networks. MobileNet applied depthwise separable convolutions to construct light-weight deep neural network, and is employed for mobile and embedded vision applications [29]. Other notable light-weight networks include ShuffleNet [30], DenseNet [31] and its follow-up work CondenseNet [32].

Classical U-Net [33] has become a standard method for image segmentation. U-Net adopts an encoder-decoder structure extracting features and mapping low-resolution features to high-resolution space with a skip connection fusing multi-features to enhance segmentation detail. Recently, some works have attempted to use attention mechanism for U-Net architecture in image segmentation and so far promoted more precise segmentation [2]. The goal of AG is to obtain detailed information of the target while suppressing

useless information. Therefore, in the last few years, attention mechanism has been added in different deep neural networks for various tasks, including ScleraSegNet [34], Attention Dense-U-Net [35], etc.

III. FOREST FIRE IMAGE DATASETS

For the purpose of experiment, we established two new datasets for segmentation and classification respectively based on four widely used public databases Corsican,[1] Foggia [10], Cair[2] and Bilkent.[3] Corsican fire dataset contains 1135 fire images captured under different environments, all images have been segmented manually as ground truth. This fire dataset can be downloaded for research purposes via a customized interface. Foggia is a video dataset consisting of 14 fire and 17 non-fire ones. Cair dataset contains normal images and images with fire from a variety of scenarios and different fire situations, such as intensity, luminosity, size, environment, etc. The Bilkent dataset is a collection of 40 video clips including 13 fire videos that can be accessed publicly and has been commonly applied to benchmark fire detection framework.

Dataset I: This dataset is used for segmentation experiments, therefore we select some fire images provided with ground truth among all datasets. In this dataset, a total of 1135 fire images from Corsican fire dataset are engaged, and 5000 images are randomly selected from the 17 non-fire videos of Foggia. In general, the 6135 images for segmentation adopted here are divided into training and testing sets by ratio 7:3.

Dataset II: This dataset is used for classification experiments. A collection of 3690 training images including 110 fire images and 110 non-fire images from Cair dataset, 455 fire images and 1880 non-fire images from Foggia dataset are engaged. For testing phase, 1631 input images are randomly selected from 40 Bilkent videos by resampling using linear interpolation with a resolution of 224×224 pixels.
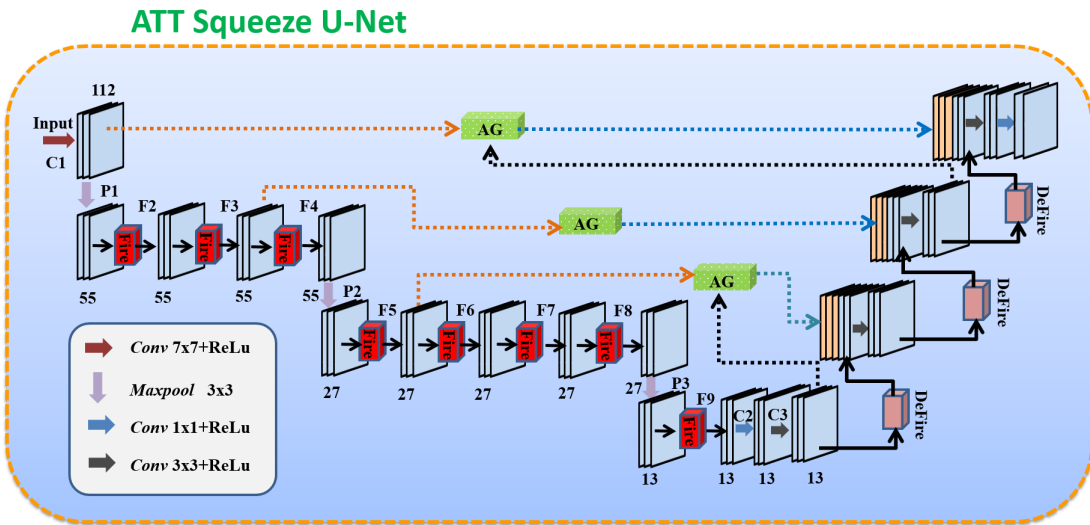
IV. METHODOLOGY
A. PROPOSED NETWORK FRAMEWORK

Many researches attempted to achieve higher detection accuracy by increasing the depth of network. However, deep CNN architecture requires a large number of network parameters, which are not suitable for fire detection and other real-time applications. In addition, another barrier of establishing CNN-based architecture is the large number of training data required. However, accessing forest fire images for fire detection is even more challenging than common tasks. Inspired by these two analyses, squeeze architecture that only require a few model parameters and U-Net with attention mechanism that could highlight regions-of-interest (ROI) while suppressing irrelevant features could be regarded as good candidates for real-time fire detection. Therefore, this study incorporates SqueezeNet structure into Attention

---

[1]http://cfdb.univ-corse.fr/
[2]https://github.com/UIA-CAIR/Fire-Detection-Image-Dataset
[3]http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips.html

FIGURE 1. This is the architecture of our proposed ATT Squeeze U-Net. There is a contracting path which uses a SqueezeNet architecture with eight modified Fire modules, and an expansion path incorporating three DeFires modules that take the same ideology as our modified Fire module. Three AGs at the skip connection concatenate encoder and decoder where the dotted arrows in three colors represent three signals. The solid black arrows represent the route of feature maps.

U-Net architecture [2] to enhance local feature expression and improve the performance of fire detection. The architecture of ATT Squeeze U-Net is shown in Fig. 1.

1) MODIFIED SQUEEZENET WITH NOVEL FIRE MODULE

Classical SqueezeNet [3] is a less complex model which is composed of a basic structure called Fire module. Small convolution kernels have been used to reduce parameter size and memory demand while accuracy maximized. Conv1 is the first layer of SqueezeNet, followed by eight Fire modules and ended with a final convolution layer. Fire module is the main block of SqueezeNet, and is composed of a squeeze layer and an expand layer that has a concatenation of 1×1 and 3×3 filters. However, the 3×3 filters generate a large number of parameters with only basic functions are played through the whole Fire module.

Based on the above analysis, we design a new Fire module that could reduce the number of parameters and increase learning ability effectively. Fig. 2 shows the structure of this modified Fire module, where the 1×1 kernel that trains simultaneously with the original 3×3 filter still remain as conventional Fire module design. However, the original 3×3 filters in the expand layer are replaced here by an 3×3 DWConv kernel and a Channel Shuffle operation due to the following two reasons: (i) The depthwise separable convolution is a form of divided convolution which divides a standard convolution into a depthwise convolution and a pointwise convolution of 1×1 kernel size [29]. The use of depthwise separable convolution can reduce network complexity with fair precision maintained by keeping the number of training weight parameters needed at a low level. Depthwise separable convolution can make it possible to separate channels from convolution region, as well as connect input and output feature maps one-to-one through convolution operation; (ii) the Channel Shuffle [30] can solve the problem that the output
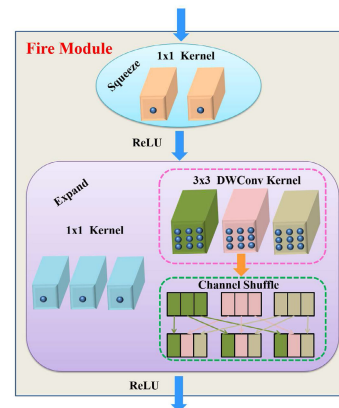
FIGURE 2. The structure of the proposed Fire module.

of a channel is only related to its corresponding input, so that information exchange between channels and feature description could be enhanced. It divides the channels in each group into several sub-groups and feed each group with different subgroups in the next layer. In this way, Channel Shuffle ensures information exchange between different groups of channels and improves the accuracy of feature description.

2) ATT SQUEEZE U-NET

In this study, we remove the encoder of Attention U-Net by our modified SqueezeNet for more effective feature extracting. In order to match channel numbers of our SqueezeNet with initial U-Net layer, channel numbers in the first convolution layer of the modified SqueezeNet is changed from 96 to 64. After that, the prediction and classification functions of SqueezeNet realized by average pooling layer and softmax classifier at the end could be abandoned while embedding in our Attention U-Net. Finally, we add another 3×3 convolution layer between modified SqueezeNet and the decoder to
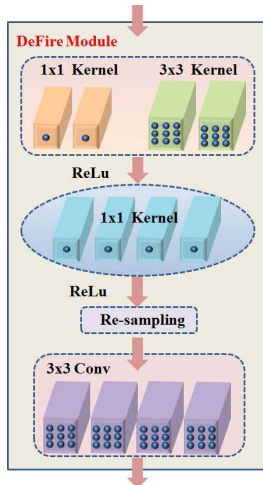
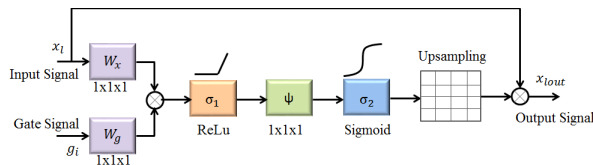**FIGURE 3.** The architecture of the proposed DeFire module.

FIGURE 4. Schematic of additive Attention Gate.

improve extracting accuracy. By these designs, the replacement of SqueezeNet enables only limited calculation and storage space required for the proposed ATT Squeeze U-Net. Also, the DWConv replaced and Channel Shuffle added in Fire module could improve feature communication and reduce computational cost. The whole encoder of Attention U-Net (a contracting path on the left side) is shown in Fig. 1.

After modifying Fire module in SqueezeNet by two aspects and integrating it into the encoder of Attention U-Net, a corresponding DeFire module and an AG connection is also adopted in the whole network architecture as shown in Fig. 1. Reverse to the Fire modules incorporated in encoding path, another DeFire module would be designed to replace traditional up-sampling step of Attention U-Net decoder as shown in Fig. 3. The proposed DeFire module consists of an extend layer and a squeeze layer, in which the extend layer is built of 1×1 and 3×3 convolutional filters and the squeeze layer is made of a 1×1 convolutional filter, a re-sampling layer and a 3×3 convolutional layer. In this way, we can expand feature maps in an equivalent efficient way as the encoding Fire module using DeFire module.

The use of AG [2] as shown in Fig. 4 in skip connection also contribute to more focusing information transmission. In this figure, x_l denotes feature maps generated by the encoder from current scale, and g_l is gating signal collected from a coarser scale. AG takes x_l and g_i as the input of attention to achieve attention coefficient α

$$\alpha = \sigma_2(\psi^T(\sigma_1(W_x^T x_l + W_g^T g_i + b_g)) + b_\psi), \qquad (1)$$

where σ_1 is often chosen as ReLU function, e.g. $\sigma_1(x) = \max(0, x)$, and σ_2 adopts a Sigmoid function defined as
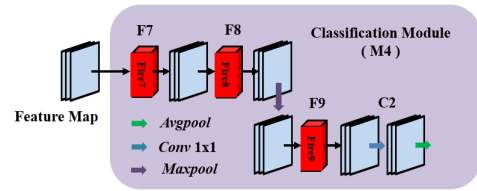

FIGURE 5. The classification structure of training module.

$\sigma_2(x) = 1/(1 + e^{-x})$. W_x, W_g and ψ are weights of linear transformations in the form of channel-wise 1×1x1 convolutions, while b_g and b_ψ are bias terms of linear transformations. The output of AG is the element-wise multiplication of x_l and α.

$$x_{lout} = x_l \times \alpha. \qquad (2)$$

Finally, the results of the last DeFire module is then fed into 3×3 convolution followed by a final 1×1 convolutional layer. The last 1×1 convolution layer is used to map 64-channel feature vector to the desired number of classes to predict each pixel.

3) M4 FOR CLASSIFICATION

Since the segmentation module could only segment areas suspected as fire, while the incorrectness of identifying fire from similar objects still remains. Therefore, we adopt an advanced classification technique to identify real fire images from network predicted results. The encoding fragment of our ATT Squeeze U-Net can be regarded as an efficient classifier of fire image recognition task. After a general consideration of the effectiveness and number of parameters, this article adopts the classification architecture named M4 shown in Fig. 5. Specifically, all feature maps after Fire module (F6) with size of $27 \times 27 \times 384$ are fed to the following two Fire modules, followed by a 3×3 maxpooling layer. After a Fire module is applied again, the result would be convoluted with a 1×1 kernel. Finally, the avgpool of 13×13 is applied. The corresponding information of this is tabulated in Table 1, and 131470 parameters are needed. Discussions on the selection of classification modules are provided in experiment subsection.

B. FOREST FIRE DETECTION AND RECOGNITION

The proposed approach attempts to segment all fire like areas, and then identify whether they are fire or not by the classification module. Fig. 6 presents the schematic flow diagram of the training module, which consists of three major steps. In the first step, input images from Dataset I are fed into ATT Squeeze U-Net for training. The obtained segmentation model would then be trained by Dataset 2 for achieving feature maps in step 2. Then the feature map would be selected to train our classification module M4. Finally, a softmax layer for classification [29] set after fully connected layer is used to generate category probability of fire existence.

Fig. 7 shows the testing process on this whole proposed work. In actual testing process, images would first be delivered into segmentation and classification procedures as Fig. 6

TABLE 1. The architecture of our classification module (M4).

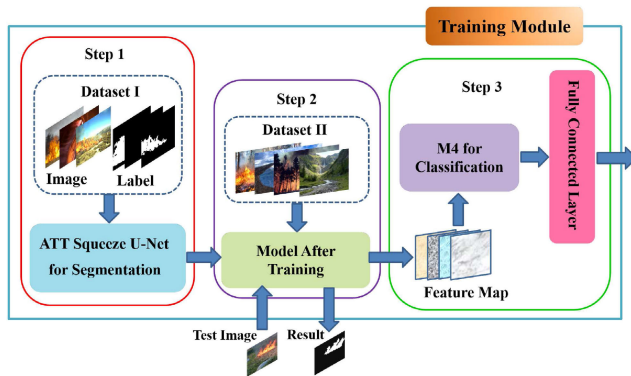| Layers	Output Size	Kernel Size/stride	$s_{1\times1}$ conv (1×1 Squeeze)	$e_{1\times1}$ conv (1×1 Expand)	$e_{3\times3}$ DWConv (3×3 Expand)
Input	27 × 27 × 384				
F7	27 × 27 × 384		48	192	192
F8	27 × 27 × 512		64	256	256
Maxpool	13 × 13 × 512	3 × 3/2			
F9	13 × 13 × 512		64	256	256
C2	13 × 13 × 2	1 × 1/1			
Avgpool	1 × 1 × 2	13 × 13/1			

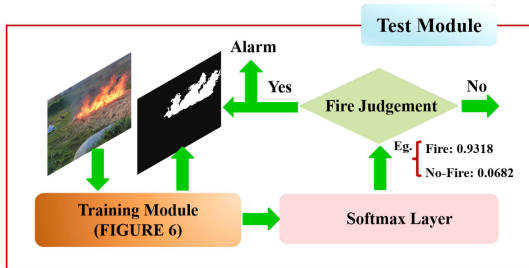**FIGURE 6.** The architecture of the proposed training module.

FIGURE 7. The architecture of the proposed test module.

inputting from the "Model After Training" module. Some testing results operated by the segmentation model in Step 2 would be illustrated in Fig. 8. Having passed through all remaining procedures in the "Training Module", some following processes for results presentation are continued as shown in Fig. 7. Then, category probability by softmax layer of every test image would be conveyed into "Fire Judgement" module. Classification results with category probabilities larger than 0.8 would be identified as fire existed and have their segmentation results output with an alarm. Those segmentation images with classification results as fire absent would be abandoned, continuing with the next test image.

V. EXPERIMENTAL RESULTS AND ANALYSIS
A. PERFORMANCE ENVIRONMENT AND EVALUATION METRICS
The proposed model is investigated using Python 3.7 and verified over Datasets I and II. The network architecture is built based on a publicly available Tensorflow framework and implemented on a small server with Intel (R) Core (TM)

i7-7700K CPU (4.5 GHz) with 32GB memory, NVIDIA GeForce GTX 1070 (GPU) with 8GB of memory.

The performance of this proposed approach is evaluated by Sensitivity (*SE*), Specificity (*SP*), Accuracy (*ACC*), Dice Similarity Coefficient (*DSC*), Precision and Recall. They are defined as follows

$$SE = \frac{TP}{TP + FN}, \quad (3)$$

$$SP = \frac{TN}{FP + TN}, \quad (4)$$

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}, \quad (5)$$

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}, \quad (6)$$

$$Precison = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad (7)$$

where *TP*, *TN*, *FP*, *FN* denote the amount of true positive, true negative, false positive, and false negative, respectively. In addition, the performance has also been examined in terms of standard indexes, such as AUC (area under the curve) and ROC (receiver operating characteristic curve). AUC values are calculated using the trapezoidal rule. The closer the AUC value is to 1, the better performance the corresponding forest fire segmentation algorithm achieves. The ROC curve plots the change of *SE* versus 1-*SP* by varying the threshold on probability map.

B. CLASSIFICATION MODULE SELECTION
To choose the best classification module with few network parameters, this article discusses five cases of classification modules and compares the effect of different modules, which are shown in Fig. 9. In order to know which part of the encoder of our ATT Squeeze U-Net architecture would be the best performing classification modules, we combine these classification modules with the subsequent fully connected layers one by one, as shown in Fig. 6. We compare the accuracy of different classification modules when applying separately to fire detection and recognition. The results presented in accuracy metrics plotted against the number of parameters needed are shown in Fig. 10. It could be seen from the results that M4, M5 could bring the accuracy of fire identification to a considerably high level. However, adopting M4 as the classifier, relatively few network parameters are needed to achieve the similar recognition rate as M5.

FIGURE 8. Results form trained segmentation module, up: original test images, down: segmentation results using the proposed segmentation framework.
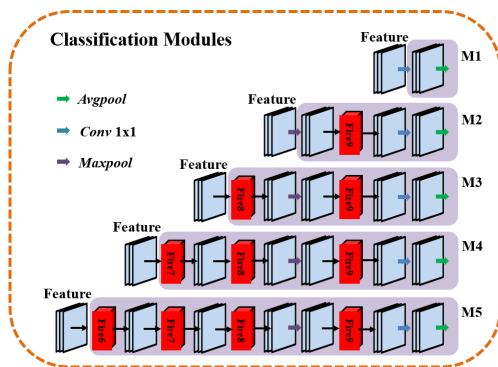

FIGURE 9. The architectures of five implementations of the classification modules, the solid black arrows represent the route of feature maps.
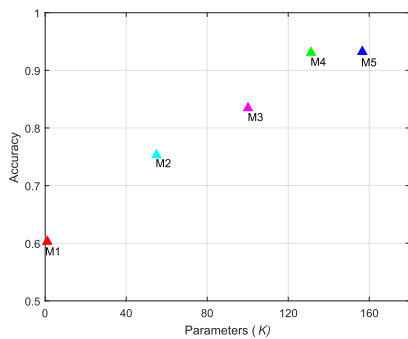

FIGURE 10. Fire recognition performance using different classification modules.

From above experiments, M4 stands out to be the best suitable classification module.

C. FIRE LOCATION AND SEGMENTATION EVALUATION

To evaluate the segmentation performance of the proposed framework, we compare it against five different deep neural network models, U-Net[4] [33], Attention U-Net[5] [2], DenseNet[6] [31], R2U-Net[7] [36], SegNet[8] [37]. The implementation of these models is based on publicly

[4]https://github.com/ternaus/robot-surgery-segmentation
[5]https://github.com/ozan-oktay/Attention-Gated-Networks
[6]https://github.com/SimJeg/FC-DenseNet
[7]https://github.com/LeeJunHyun/Image_Segmentation#2u-net
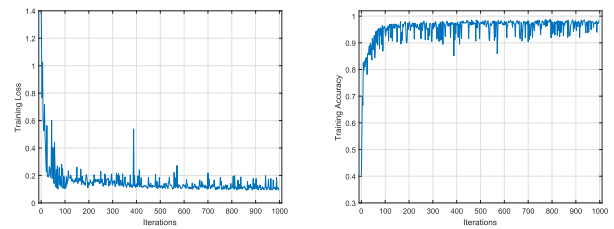[8]https://github.com/tkuanlun350/Tensorflow-SegNet

FIGURE 11. The loss curves and *ACC* of the proposed ATT Squeeze U-Net as the number of iterations increases during training stage.

available codes. In order to adapt the images to different network architectures, we adjust input images to the size required by each network. We first train the proposed segmentation framework using the training set from Dataset I, and then use the testing images of this dataset to test the trained model. The 1137 Corsican images with ground truth distributed in both training and testing sets would meet requirements for both segmentation network learning and quantitative analysis. Specifically, input images through down-sampling using linear interpolation with a resolution of 224×224 pixels are fed into the ATT Squeeze U-Net. The network parameters are optimized using the Adam optimizer [38]. Taking batch size as 12 during training and the verified batch size as 6, the initial learning rate is set to 1×10^{-4} and the learning decay rate is 0.99. The training accuracy and loss curves of ATT Squeeze U-Net in Fig. 11 shows that this model is stable and converges rapidly with the increase of the number of iterations. Fig. 11 demonstrates a quick convergence and a flat curve for the rest of the iteration, which indicates the stability throughout training and the efficient obtainment of optimal solution. Therefore, a relatively high reliability as a novel proposed network aiming at achieving real-time fire detection is shown.

Fig. 12 shows the qualitative comparisons of ATT Squeeze U-Net with other deep networks. The qualitative results show good capability of our model in extracting fire details. R2U-Net shows superior performance as compared to SegNet and DenseNet. After observing the segmented fire areas, Attention U-Net performs better than traditional U-Net model obviously.

To validate the introduced scheme, we compared the performance of ATT Squeeze U-Net with other existing networks. All predicted results using the testing set of Dataset I

FIGURE 12. Visual comparison of various fire detection modules. From left to right: original fire images, ground truth, DenseNet, Attention U-Net, U-Net, SegNet, R2U-Net, and Ours, respectively.

TABLE 2. Comparison results of various deep convolution models for fire extraction on Dataset I, highlighted value represents the best result.

Models	Parameters (M)	SE	SP	ACC	DSC
DenseNet	8.19	0.8432	0.9214	0.8918	0.8082
Attention U-Net	8.42	0.7958	0.9262	0.8903	0.7986
U-Net	**7.76**	0.7163	0.9093	0.8852	0.8100
SegNet	14.52	0.7483	0.9078	0.8887	0.8243
R2U-Net	8.23	**0.8692**	0.9183	0.9002	0.8613
Ours	7.96	0.8205	**0.9273**	**0.9067**	**0.8750**

are compared with ground truth, and final evaluation scores are listed in Table 2. As can be seen from this table, the *SE* and *SP* values of R2U-Net both stay at a relatively good level, whereas our model is slightly more specific than the other five methods. R2U-Net however is slightly more sensitive at capturing any suspicious fire region. This is mainly due to the advanced feature accumulation process by recurrent residual layers that enable better representation for segmentation. The attention mechanism contributes to higher performance of Attention U-Net compared to classical U-Net apparently. Both U-Net and SegNet show generally less competitive performance than the other four networks. At the same time, it could be seen that the *SP* of our model is more superior than *SE*, while *ACC* and *DSC* of the proposed framework are the highest among all discussed models. Our figures are within a relatively high position among all models, as well as a notable increase especially on *SE* and *DSC* compared to classical Attention U-Net is seen by the incorporation of our modified Fire module in SqueezeNet. Besides, model parameters are

also measured to testify computational complexity. It could be seen that our method uses 7.96M parameters, slightly less than 8.42M of Attention U-net. This is mainly due to the substituted SqueezeNet fragment significantly reduce parameters than original Attention U-Net. For comprehensive consideration on segmentation results and computational cost, this exceeding parameter number is reserved for trade-off. However, our method is still relatively less complicated compared to 8M parameters on average of most experiment approaches and SegNet, where more than 14M parameters are engaged. In general, demonstrate the effectiveness of the proposed lightweight network on forest fire location and extraction.

For a more comprehensive validation, the ROC curve is also introduced to evaluate fire extraction of our network. In this experiment, the ROC plots *SE* against 1-*SP*, and the corresponding average values of area under ROC curves for Dataset I are illustrated in Fig. 13. AUC evaluates the balance of fire segmentation when *SE* arrived at the detriment of *SP*. It can be seen that the ROC curve of our model is notably closer to the upper left corner. Fig. 13 also displays the precision evaluation for ROC curve and AUC values on testing images for forest segmentation.

D. FIRE RECOGNITION EVALUATION

Feeding training images of Dataset II into ''Model after Training'' module shown as Step 2 of Fig. 6, all feature maps could be obtained. Then, the proposed classification module
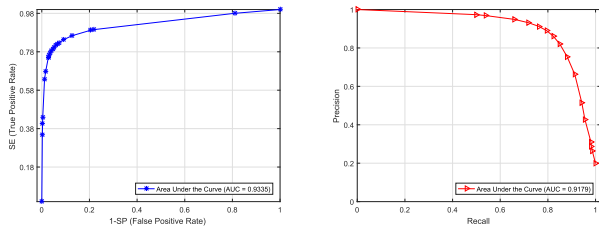
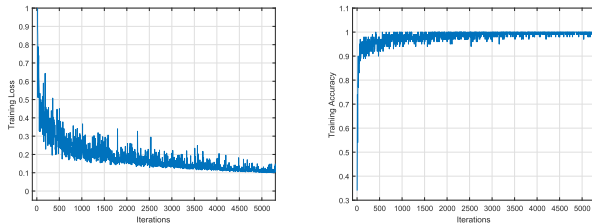**FIGURE 13.** ROC curve analysis of the proposed network for fire extraction.

FIGURE 14. The loss and *ACC* curves of the proposed classification network (see Step 3 in Fig. 6) as the number of iterations increases during training process.

(M4) is trained on these feature maps using Adam optimizer. We set the parameters during training to a specific value, batch size as 16, initial learning rate as 0.0001, and learning decay rate as 0.99. Fig. 14 shows that the classification network can converge quickly in training process, and the loss value rapidly stay controlled. We can also observe from the accuracy curve from Fig. 14 that our method achieves convergence performance approaches to one. Fig. 15 shows some examples of the prediction score of our recognition model. It can be seen that our algorithm can correctly distinguish fire and non-fire images. It can be seen that category probabilities show a relatively certain degree of classification, and the dominant scores match with actual existence of fire. Sunset images are also classified as Non-fire even though a lower score of Non-fire probability appears compared to other non-fire images. To exam the robustness of the proposed

TABLE 3. Fire recognition results using the proposed framework on testing set of Dataset II.

Video Clips		False Frames	True Frames	Accuracy
Video1		12	196	94.23%
Video2		63	1138	94.75%
Video3		5	203	97.60%
Video4		10	206	95.37%
Video5		12	128	91.43%
Video6		19	241	92.69%

M4 model, we use the testing set from Dataset 2 for classification. Since the fire varies slowly in video frames, successive frames are highly similar. To reduce computation cost, we use one frame every 20 frames from the video clip for processing. Table 3 lists some fire detection results of our framework in six video scenes. It can be noted from this table that our method achieves qualified accuracy in recognizing forest fire.

To further evaluate the recognition performance of our network model, we compare it against four different deep convolutional networks SqueezeNet[9] [3], ResNet 50[10] [39], VGG 16[11] [25], BLS[12] [40], MobileNet-Fire [41], EMN-Fire [28] and Muhammad *et al.* [17]. The testing set from Dataset II is used to compare all these models, and the results in terms of three evaluation metrics *FN*, *FP*, and *ACC* are listed in Table 4. Table 4 shows that deep convolutional models achieve generally superior value on *ACC*. This is probably because convolutional neural network-based frameworks

[9]https://github.com/DeepScale/SqueezeNet
[10]https://github.com/liangyihuai/my_tensorflow/tree/master/com/huai/converlution/resnets
[11]https://github.com/dhuQChen/VGG16
[12]http://broadlearning.ai/download-bls/

(a) Fire: 97.52%, (b) Fire: 98.02%, (c) Fire: 95.88%, (d) Fire: 96.34%, (e) Fire: 91.14%, (f) Fire: 93.36%,
Non-fire: 2.48%; Non-fire: 1.98%; Non-fire: 4.12%; Non-fire: 3.66%; Non-fire: 8.86%; Non-fire: 6.64%;
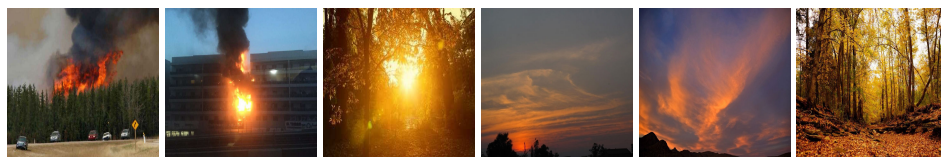
(g) Fire: 95.93%, (h) Fire: 93.26%, (i) Fire: 13.16%, (j) Fire: 9.170%, (k) Fire: 10.86%, (l) Fire: 7.680%,
Non-fire: 4.07%; Non-fire: 6.74%; Non-fire: 86.84% Non-fire: 90.83% Non-fire: 89.14% Non-fire: 92.32%

FIGURE 15. Category probability results of our fire classification model.

TABLE 4. Comparison of our method on Dataset II using different metrics, highlighted value represents the best result.

| Models	Parameters	FN	FP	ACC
SqueezeNet [3]	1.25×10^6	0.62	7.87	0.9301
ResNet 50 [39]	2.55×10^7	**0.00**	9.57	0.9044
VGG 16 [25]	1.38×10^8	1.20	10.01	0.9001
BLS [40]	$\mathbf{1.17 \times 10^5}$	2.90	7.32	0.9025
MobileNet-Fire [41]	1.40×10^6	3.98	4.49	0.9317
EMN-Fire [28]	3.40×10^6	4.68	4.29	0.9279
Muhammad et al. [17]	6.80×10^6	4.74	4.00	0.9193
Ours	1.21×10^5	4.58	**3.91**	**0.9321**

obtaining features in universal pattern instead of selecting manually. In addition, the *FN* metric of ResNet50 shows the best *FN* rates at 0, which indicates high reliability of ResNet50 if results shown as non-fire. However, ResNet50 is of higher possibility raising mistaken fire recognition as indicated by its high *FP* (9.57). Similarly, SqueezeNet, VGG 16 and BLS all seem to be too sensitive to fire-like images, which generally high *FP* results in their decreasing *ACC*. MobileNet-Fire tends to have lower *FN* ratio (3.98) than ours, which is mainly due to the advanced channel multiplier network architecture that is highly sensitive to specific color. Therefore, *FP* ratio for MobileNet-Fire is higher due to their color-based mechanism as well (4.49). MobileNet-Fire, EMN-Fire and network proposed in [17] all achieve relatively competitive *ACC* and overall lower *FN* and *FP*, however parameters engaged are tenfold more than ours. As a lightweight network, our approach employs a total of 120768 parameters second only to 116872 by BLS. BLS engage less computation than other CNN networks mainly due to its broad structure, but a slightly lower *ACC* is achieved by BLS at 0.9025. Our model shows a relatively better trade-off by achieving the highest accuracy at 0.9321 and reducing computational cost based on recent advanced architecture. This is mainly because of the DWConv replaced in Fire module that decreases computation while preserving good feature learning ability. The channel shuffle operation added in Fire module increases inter-channel feature communication that would bring higher accuracy in general.

We also evaluate all approaches through a true positive fraction (*SE*) verses false positive fraction (1-*SP*) point diagram as shown in Fig. 16. Methods as close as to point (0, 1) are regarded as comprehensively advanced among networks. It could be seen that latest recognition networks [17], [41] [28] seem to enhance *SE* and *SP* to a large extent. Compared to Mobile-Net, EMN-Fire and framework proposed by Muhammad *et al.* [17], our network tends to bring improvement in terms of specificity.

E. SEGMENTATION AND RECOGNITION FOR SMALL FIRE
In order to evaluate the performance of our approach on tiny fire detection, several images with insignificant fire are selected for a further quantitative analysis. Visual comparisons with other approaches as shown in Fig. 17. It could be seen that while Attention U-Net, U-Net and SegNet obtain segmentation contours very different from ground truth,
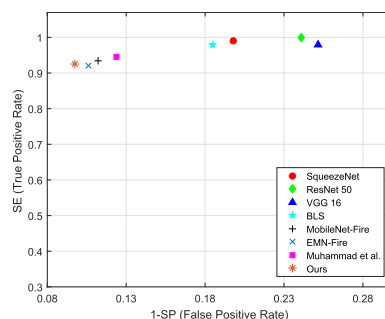
FIGURE 16. Comparison of various fire classification models.

TABLE 5. Quantitative segmentation comparison for tiny fire images from Dataset I, highlighted value represents the best result.

Models	SE	SP	ACC	DSC
DenseNet	**0.8137**	0.8501	0.8358	0.7764
Attention U-Net	0.7555	0.8615	0.8302	0.7698
U-Net	0.6767	0.8393	0.8060	0.6936
SegNet	0.7183	0.8078	0.8159	0.6843
R2U-Net	0.8001	0.8237	0.8269	0.7721
Ours	0.7731	**0.8611**	**0.8582**	**0.7866**

DenseNet, R2U-Net and our approach gain relatively reliable results. However, our approach is slightly more competitive in terms of fire shapes and details. Recognition results of our approach are shown in Fig. 18, where a reliable probability of fire existed is detected under tiny fire circumstance.

Quantitative results in Table 5 show that our approach is relatively sensitive and is of high specificity while detecting inapparent fire. Besides, we achieve the highest *ACC* and *DSC* among all comparison methods at 0.8582 and 0.7866 respectively. Similar to our qualitative comparison, DenseNet and R2U-Net also gain reliable accuracy. In terms of classification performance listed in Table 6, our approach shows a relatively good trade-off between *FN* and *FP* at 5.01 and 6.01. In general, all methods present a slightly higher false detection rate in tiny fire detection. Although SqueezeNet, ResNet 50 and VGG 16 have *FN* controlled within 1, their *FP* increase to about 10. Recent studies [28], [41] [17] are relatively prudent on small fire detection, which a slightly higher *FP* ratio than us is obtained. Due to the advanced structure proposed by MobileNet-Fire [41], *FN* is maintained at the lowest level among all methods which is 0.06 less than ours. Similar to previous quantitative comparisons, our approach shows a comparably better tradeoff between *FN* and *FP* and obtain a slightly better *ACC* at 92.07. Although accuracy for tiny fire detection decrease for about 0.01, overall recognition for the network still remains acceptable.

F. DISCUSSION OF TRAINING TIME
From the calculation point of view, less calculating time during training indicates better effectiveness of the model. Therefore, in this experiment, we report the training and prediction time of segmentation and recognition processes.
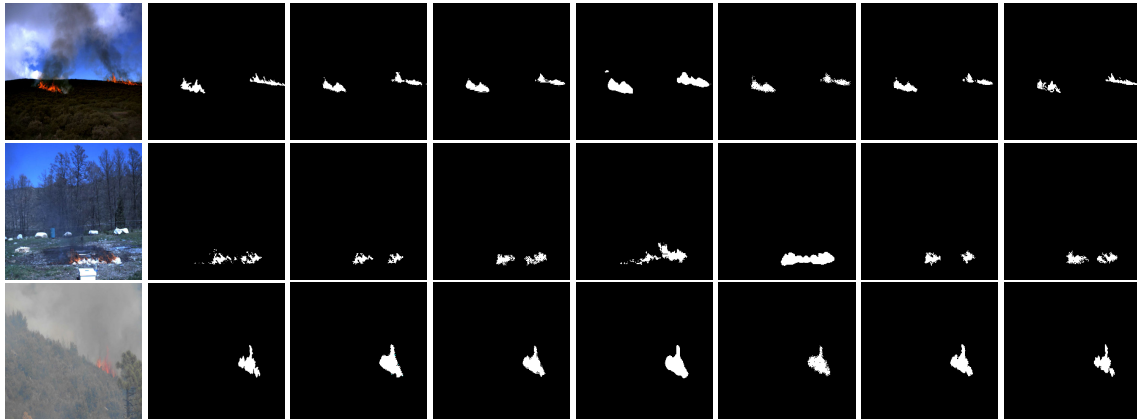
FIGURE 17. Qualitative comparison for tiny fire segmentation. From left to right: original fire images, ground truth, DenseNet, Attention U-Net, U-Net, SegNet, R2U-Net, and Ours, respectively.

TABLE 6. Quantitative recognition comparison for tiny fire images from Dataset I, highlighted value represents the best result.

Models	FN	FP	ACC
SqueezeNet [3]	0.71	8.92	0.9186
ResNet 50 [39]	**0.13**	10.17	0.9022
VGG 16 [25]	1.18	9.65	0.9153
BLS [40]	2.75	9.24	0.8926
MobileNet-Fire [41]	4.95	6.72	0.9196
EMN-Fire [28]	5.17	6.25	0.9191
Muhammad et al. [17]	5.23	6.12	0.9014
Ours	5.01	**6.01**	**0.9207**

(a) Fire: 90.64%, (b) Fire: 90.08%, (c) Fire: 91.53%, Non-fire: 9.36%; Non-fire: 9.92%; Non-fire: 8.47%;

FIGURE 18. Category probability results for tiny fire using our fire classification model.

TABLE 7. Training time and prediction time of various segmentation networks on Dataset I, highlighted value represents the best result.

Methods	Training time (s)	Prediction time (s)
Attention U-Net [2]	2931.44	2.0617
U-Net [33]	**2390.38**	**1.4088**
DenseNet [31]	2729.38	1.6365
SegNet [37]	3965.87	2.3423
R2U-Net [36]	2867.23	1.9108
Ours	2638.15	1.5432

TABLE 8. Training and prediction time of various classification models on Dataset II, highlighted value represents the best result.

Methods	Training time (s)	Prediction time (s)
SqueezeNet [3]	12290.96	1.2854
ResNet 50 [39]	15833.78	1.4188
VGG 16 [25]	24213.39	1.6176
BLS [40]	**130.06**	**0.7309**
MobileNet-Fire [41]	12874.46	1.2907
EMN-Fire [28]	14792.42	1.3523
Muhammad [17]	13227.14	1.3027
Ours	400.27	0.8941

It could be seen from the operation time comparison of six models from Table 7 that U-Net obtains competitive results comparing with other models. DenseNet performed for 2729.38 seconds, which is 1.45 times faster than the slowest 3965.87 seconds of SegNet. Due to the addition of AG, it can be seen that training time of Attention U-Net is longer than U-Net. Our proposed model consumes 2638.15 seconds, which is 293.29 seconds less than Attention U-Net due to the modified SqueezeNet applied. In addition, testing time for a single image shows that there is no significant difference between each method, whereas our approach uses slightly fewer time compared to most methods at 1.54 seconds.

Table 8 shows the training time and testing time of these five classification models using Dataset II. From this table, we can see that BLS architecture generally takes less time to process a large number of training data comparing to other methods. The relatively few computational costs of BLS could be explained as the only two layers engaged in training. Our system trained for 400 seconds, which is

approximately 30 times faster than 12290 and 12874 seconds of SqueezeNet and MobileNet-Fire. In addition, VGG 16, ResNet 50 and EMN-Fire need more time-consuming deep structures to ensure their accuracies. Generally, the proposed lightweight network processes for significantly fewer time than most of the current recognition methods. In addition, Table 8 depicts that the predicting time by all models in test phase is similar, whereas our method uses slightly fewer time compared to most methods at 0.89 seconds.

VI. CONCLUSION

In this article, we firstly proposed ATT Squeeze U-Net for segmentation and recognition. The incorporated SqueezeNet architecture with modified Fire module on ATT U-Net, which enabled more effective feature learning based on limited data. Subsequently, another recognition model adopting a portion of the newly established encoding path was utilized for classification. Apart from providing existing segmentation and recognition frameworks with a more efficient alternative, this study also verified its effectiveness on fire recognition where
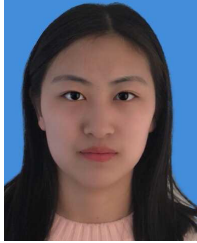
high sensitivity was required and limited training data could be obtained. Experiments showed that the proposed architecture achieved relatively competitive segmentation accuracy and reliable recognition. However, there might still be some limitations in terms of comprehensive fire detection even though relatively accurate fire circumstances were alarmed and fire regions could be segmented in precise detail.

Temporal factors such as flame development, fire spread and color variance may be difficult to evaluate which is also considered valuable in fire detection. Besides, various weather conditions such as foggy and snowing may also hinder network recognition. For future researches, video analysis and network modifications on segmentation and recognition might be made based on the proposed architecture. We may also focus on adjusting the proposed network architecture and fire detection under specific scenarios.

REFERENCES

[1] B. Liu, D. Zou, L. Feng, S. Feng, P. Fu, and J. Li, "An FPGA-based CNN accelerator integrating depthwise separable convolution," *Electronics*, vol. 8, no. 3, pp. 281–282, 2019.

[2] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: http://arxiv.org/abs/1804.03999

[3] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360

[4] J. Li, B. Yan, M. Zhang, J. Zhang, B. Jin, Y. Wang, and D. Wang, "Long-range Raman distributed fiber temperature sensor with early warning model for fire detection and prevention," *IEEE Sensors J.*, vol. 19, no. 10, pp. 3711–3717, May 2019.

[5] X. Qiu, Y. Wei, N. Li, A. Guo, E. Zhang, C. Li, Y. Peng, J. Wei, and Z. Zang, "Development of an early warning fire detection system based on a laser spectroscopic carbon monoxide sensor using a 32-bit system-on-chip," *Infr. Phys. Technol.*, vol. 96, pp. 44–51, Jan. 2019.

[6] L. Hua and G. Shao, "The progress of operational forest fire monitoring with infrared remote sensing," *J. Forestry Res.*, vol. 28, no. 2, pp. 215–229, Mar. 2017.

[7] W. Krüll, R. Tobera, I. Willms, H. Essen, and N. von Wahl, "Early forest fire detection and verification using optical smoke, gas and microwave sensors," *J. Forestry Res.*, vol. 45, pp. 584–594, 2012.

[8] J. Wang and H. Wang, "Tunable fiber laser based photoacoustic gas sensor for early fire detection," *Infr. Phys. Technol.*, vol. 65, pp. 1–4, Jul. 2014.

[9] G. Marbach, M. Loepfe, and T. Brupbacher, "An image processing technique for fire detection in video images," *Fire Saf. J.*, vol. 41, no. 4, pp. 285–289, Jun. 2006.

[10] P. Foggia, A. Saggese, and M. Vento, "Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1545–1556, Sep. 2015.

[11] F. Yuan, "Video-based smoke detection with histogram sequence of LBP and LBPV pyramids," *Fire Saf. J.*, vol. 46, no. 3, pp. 132–139, Apr. 2011.

[12] W. Ye, J. Zhao, S. Wang, Y. Wang, D. Zhang, and Z. Yuan, "Dynamic texture based smoke detection using surfacelet transform and HMT model," *Fire Saf. J.*, vol. 73, pp. 91–101, Apr. 2015.

[13] F. Yuan, "A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with AdaBoost for video smoke detection," *Pattern Recognit.*, vol. 45, no. 12, pp. 4326–4336, Dec. 2012.

[14] Y. Chunyu, F. Jun, W. Jinjun, and Z. Yongming, "Video fire smoke detection using motion and color features," *Fire Technol.*, vol. 46, no. 3, pp. 651–663, Jul. 2010.

[15] M. Mueller, P. Karasev, I. Kolesov, and A. Tannenbaum, "Optical flow estimation for flame detection in videos," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2786–2797, Jul. 2013.

[16] M. Yin, C. Lang, Z. Li, S. Feng, and T. Wang, "Recurrent convolutional network for video-based smoke detection," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 237–256, Jan. 2019.

[17] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, "Convolutional neural networks based fire detection in surveillance videos," *IEEE Access*, vol. 6, pp. 18174–18183, 2018.

[18] S. Frizzi, R. Kaabi, M. Bouchouicha, J.-M. Ginoux, E. Moreau, and F. Fnaiech, "Convolutional neural network for video fire and smoke detection," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2016, pp. 877–882.

[19] G. Lin, Y. Zhang, G. Xu, and Q. Zhang, "Smoke detection on video sequences using 3D convolutional neural networks," *Fire Technol.*, vol. 55, no. 5, pp. 1827–1847, Sep. 2019.

[20] Q.-X. Zhang, G.-H. Lin, Y.-M. Zhang, G. Xu, and J.-J. Wang, "Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images," *Procedia Eng.*, vol. 211, pp. 441–446, 2018.

[21] P. Barmpoutis, K. Dimitropoulos, K. Kaza, and N. Grammalidis, "Fire detection from images using faster R-CNN and multidimensional texture analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 8301–8305.

[22] Z. Jiao, Y. Zhang, J. Xin, L. Mu, Y. Yi, H. Liu, and D. Liu, "A deep learning based forest fire detection approach using UAV and YOLOv3," in *Proc. 1st Int. Conf. Ind. Artif. Intell. (IAI)*, Jul. 2019, pp. 1–5.

[23] A. J. Dunnings and T. P. Breckon, "Experimentally defined convolutional neural network architecture variants for non-temporal real-time fire detection," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 1558–1562.

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

[25] P. Matlani and M. Shrivastava, "Hybrid deep VGG-NET convolutional classifier for video smoke detection," *Comput. Model. Eng. Sci.*, vol. 119, no. 3, pp. 427–458, 2019.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[27] Y. Peng and Y. Wang, "Real-time forest smoke detection using hand-designed features and deep learning," *Comput. Electron. Agricult.*, vol. 167, pp. 1–18, Dec. 2019.

[28] K. Muhammad, S. Khan, M. Elhoseny, S. Hassan Ahmed, and S. Wook Baik, "Efficient fire detection for uncertain surveillance environment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 3113–3122, May 2019.

[29] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[30] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[32] G. Huang, S. Liu, L. V. D. Maaten, and K. Q. Weinberger, "CondenseNet: An efficient DenseNet using learned group convolutions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2752–2761.

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[34] C. Wang, Y. Wang, Y. Liu, Z. He, R. He, and Z. Sun, "ScleraSegNet: An attention assisted U-net model for accurate sclera segmentation," *IEEE Trans. Biometrics Behav. Identity Sci.*, vol. 2, no. 1, pp. 40–54, Dec. 2020.

[35] S. Li, M. Dong, G. Du, and X. Mu, "Attention dense-U-net for automatic breast mass segmentation in digital mammogram," *IEEE Access*, vol. 7, pp. 59037–59047, 2019.

[36] M. Zahangir Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: http://arxiv.org/abs/1802.06955

[37] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[40] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.

[41] H. Yang, H. Jang, T. Kim, and B. Lee, "Non-temporal lightweight fire detection network for intelligent surveillance systems," *IEEE Access*, vol. 7, pp. 169257–169266, 2019.

PENGYU WANG received the B.S. degree in automation and M.S. degree in control engineering from Hebei University, Baoding, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Department of Electronics and Communication Engineering, East China University of Science and Technology, Shanghai, China. His research interests include image processing, deep learning, computer vision, and pattern recognition.

JIANMEI ZHANG received the B.S. degree in communication engineering from Ludong University, Shangdong, in 2018. She is currently pursuing the M.S. degree with the Department of Electronics and Communication Engineering, East China University of Science and Technology, Shanghai, China. Her research interests are image processing, deep learning, computer vision, and pattern recognition.

HONGQING ZHU (Member, IEEE) received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2000. From 2003 to 2005, she was a Postdoctoral Fellow with the Department of Biology and Medical Engineering, Southeast University, Nanjing, China. She is currently a Professor with the East China University of Science and Technology, Shanghai. Her current research interests include medical image processing, deep learning, computer vision, and pattern recognition. She is a member the IEICE.

XIAOFENG LING (Member, IEEE) received the B.S. and Ph.D. degrees from Shanghai Jiao Tong University, China, in 2006 and 2012, respectively. From 2013 to 2015, he served as the Director of research and development in a start-up company and committed to the research and development of wireless communication technology. He is currently a Lecturer with the School of Information Science and Engineering, East China University of Science and Technology. His research interests include image communication, deep learning, computer vision, and signal processing.

• • •