IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 March 2023, accepted 20 March 2023, date of publication 28 March 2023, date of current version 5 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3262701

RESEARCH ARTICLE

A Deep Learning-Based Experiment on Forest Wildfire Detection in Machine Vision Course

LIDONG WANG[1,2], (Member, IEEE), HUIXI ZHANG[3], YIN ZHANG[2], KEYONG HU[1], (Member, IEEE), AND KANG AN[1], (Member, IEEE)

[1]School of Engineering, Hangzhou Normal University, Hangzhou 310018, China
[2]College of Computer Science and Technology, Zhejiang University, Hangzhou 310012, China
[3]School of Information Science and Technology, Hangzhou Normal University, Hangzhou 310018, China

Corresponding authors: Lidong Wang (wld@hznu.edu.cn), Yin Zhang (yinzh@zju.edu.cn), and Kang An (Q0070031@huqc.edu.cn)

This work was supported in part by the Zhejiang Provincial High-Education Teaching Reform Project under Grant jg20220770, in part by the Zhejiang Provincial Educational Science Planning Project under Grant 2022SCG022, in part by the AI Micro-Major Project under Grant QJWZY0201, in part by the China Knowledge Centre for Engineering Sciences and Technology (CKCEST), and in part by the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China under Grant LHY21E090004.

ABSTRACT As an interdisciplinary course, Machine Vision combines AI and digital image processing methods. This paper develops a comprehensive experiment on forest wildfire detection that organically integrates digital image processing, machine learning and deep learning technologies. Although the research on wildfire detection has made great progress, many experiments are not suitable for students to operate. Also, the detection with high accuracy is still a big challenge. In this paper, we divide the task of forest wildfire detection into two modules, which are wildfire image classification and wildfire region detection. We propose a novel wildfire image classification algorithm based on Reduce-VGGnet, and a wildfire region detection algorithm based on the optimized CNN with the combination of spatial and temporal features. The experimental results show that the proposed Reduce-VGGNet model can reach 91.20% in accuracy, and the optimized CNN model with the combination of spatial and temporal features can reach 97.35% in accuracy. Our framework is a novel way to combine research and teaching. It can achieve good detection performance and can be used as a comprehensive experiment for Machine Vision course, which can provide the support for talent cultivation in machine vision area.

INDEX TERMS Machine vision, computer science education, wildfire detection, comprehensive experiment, CNN.

I. INTRODUCTION

With the rapid development of computer technology and the popularity of cameras, machine vision technology based on artificial intelligence (AI) and digital image processing has been applied to increasing fields, such as face detection [1], wildfire detection [2], object measurement [3] and surface defect detection [4]. As an interdisciplinary course, Machine Vision combines AI and digital image processing. With the development of AI, machine vision can replace human beings with intelligent programs for some automated operations and measurements [5]. A complete machine vision system includes a camera and an image processing device. The camera firstly obtains the images, then we can recognize the target object through the computer's visual recognition algorithm, and finally the image processing device can output the target recognition result through the terminal [3]. At present, machine vision has become one of the essential skills of image and video processing practitioners, and is also an important professional course in intelligent manufacturing, computer science and technology, and other majors. With the rapid development of AI in recent years, there is an increasing demand for talents in two main application fields, natural language processing and digital image processing.

In recent years, experts at home and abroad have been exploring the reform of Machine Vision course. For example, Min and Lu [6] focused on the production practice

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

and proposed multimedia teaching and guided interactive teaching. They also suggested that experiments should not only be closely related to classroom teaching, but also in accordance with practical application needs, and can arouse students' interest. Wang et al. [7], aiming at the principle and application of machine vision in the postgraduate curriculum, integrated scientific researches, teaching and practical projects into the classroom. In this way, students can associate the project research with the development of products. Shao et al. [8] designed cocoon sorting in the field of machine vision in order to cultivate intelligent manufacturing talents under the background of new engineering subjects. Han and Liu [9] designed a machine vision experiment platform with multiple modules using Tensorflow and Opencv library to solve the problems of insufficient experiments related to machine vision, unreasonable experimental design and lack of practical data. The reform of Machine Vision course in foreign countries focuses on social awareness education and the reform of teaching methods of basic technology. For example, Sigut et al. [10] believed that the teaching theme of machine vision depends on the use of new technologies. To enable students to better understand the concept, this paper developed an application for Android operating system that can perform real-time presentation of Opencv image processing technology to help students better understand the concepts related to image processing. Cote and Albu [11] advocated integrating the social awareness module into the Machine Vision course, so as to study the social impact of technology and the technology itself. Spurlock and Duvall [12], in order to expand the educational audience of machine vision, that is, not limited to postgraduates or doctoral students, increased the development of practical cases in the field of machine vision applications and reduced the derivation of mathematical formulas to better adapt to undergraduate teaching. From the above reform research, it can be found that most teaching methods focus on the combination of theory and practice, while in practice, they focus on how to design experiments that have industrial practicality and can arouse students' interests.

Machine Vision is one of the courses that closely link theory with practice. However, at present, most universities' comprehensive experiments for undergraduates/postgraduates have problems such as outdated design, lack of practicality, and most of them only use traditional machine learning for experiments [13], [14]. Although the research on wildfire detection has made great progress, detection with high accuracy is still a big challenge. In order to solve the above problems, this paper designs an automatic forest wildfire detection framework that can also be used as a comprehensive experiment for Computer Vision course. This framework uses image processing, machine learning, and deep learning technology to achieve automatic detection and annotation of forest wildfire regions, which is a novel way to combine research and teaching. To the best of our knowledge, no previous work has explored the preceding problems to such extent. The main contributions about this paper can be summarized as follows:

(1) We propose a novel wildfire image classification algorithm based on Reduce-VGGnet, which can reduce the training parameters of VGGnet and achieve 91.20% in accuracy.

(2) We propose a novel wildfire region detection algorithm based on the optimized CNN with the combination of spatial and temporal features. The experimental results on FLAME dataset show the effectiveness of our method.

(3) We combine wildfire image classification module and wildfire region detection module to be a comprehensive experiment for Computer Vision course. It is a novel way to effectively combine teaching and scientific research, and incorporates teachers' research into the teaching process.

The rest of our paper is structured as follows. In Section II, we describe the background of the task of wildfire region detection. The design of our framework is presented in Section III. The results and analysis are presented in Section IV. Finally, the conclusion is provided in Section V.

II. EXPERIMENT REQUIREMENTS AND BACKGROUND

A. EXPERIMENT REQUIREMENTS

The comprehensive experiment designed in this paper can be applied to undergraduate/postgraduate Machine Vision courses, so that students can understand the popular technology of machine vision, and cultivate the students' ability to develop some novel algorithms on image processing and automatic recognition. Students need to understand the following knowledge: ① Digital image processing. For example, the video reading technology is used to read the image frame, the image enhancement technology is used to achieve image contrast expansion, and the image segmentation technology is used to achieve the target object segmentation. ② Machine learning. It includes basic principles and experimental evaluation of traditional Machine learning algorithms, such as SVM (Support Vector Machine) and Adaboost, etc. ③ Deep learning. It includes the fundamentals, principles and popular deep network models of deep learning, such as VGGNet (Visual Geometry Group Network), CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network).

B. EXPERIMENT BACKGROUND

The occurrence of forest wildfires often causes great damage to national economic property. In recent years, the number of trees near the electricity transmission areas has increased dramatically. In addition, extreme weather such as drought has greatly increased the frequency and intensity of forest fires [15]. If the machine vision technology can be used to monitor the wildfire in real time, it can effectively avoid the serious damage caused by the fire spreading, thus effectively reducing the direct loss caused thereby.

Because the efficiency of manual fire detection is extremely low, an increasing number of researchers use modern means to detect fire, such as UAV-IoT [16], satellite remote sensing, wireless sensor, and image fire detectors.

In recent years, with the improvement of image processing, wildfires can be detected automatically by on-site video collected by the camera installed on the line tower, which greatly improves the accuracy of wildfire detection.

Forest wildfires can be detected through smoke and flame [17], [18]. The smoke-based detection is difficult to effectively distinguish the smoke from other non-dangerous smoke such as smoke from cooking and industrial chimneys, etc., so it is not as good as the flame-based detection. Therefore, the comprehensive experiment designed in our manuscript is mainly to detect wildfires by flames.

C. RELATED WORKS

There are two ways of flame detection: static and dynamic flame detection. Static flame detection aims at a single image and detects the flame region through image processing and machine learning. The dynamic flame detection aims at a video image sequence, which uses the static spatial features and time sequence information of the images to detect the fire target. The current research about these two detection methods is introduced below.

(1) Static flame detection. It often detects the flame by extracting the color and texture features of the image. For example, Jia and Jiong [19] proposed a method combining saturation and Otsu threshold segmentation for flame detection, which was judged based on SVM by combining the shape features of the flame area with LBP texture analysis. Tan et al. [20] used RGB and HSI dual color spaces to detect flame objects. Hossain et al. [21] proposed a novel algorithm that was capable of detecting both flame and smoke from a single image using block-based color features and texture features. The static detection methods mainly rely on color, texture, or shape features, but there are background noise and interference signals in many images, such as the sun and sunset glow. Therefore, these features cannot effectively detect the flame objects, and the detection accuracy is unsatisfactory.

(2) Dynamic flame detection. This method combines the temporal information of video with the static features of the image for detection. Schultze et al. [22] proposed to use spectrogram and acoustic spectrogram to obtain flame features according to flame flicker and movement direction. This method can also monitor the movement direction of flame. Xie et al. [23] used dynamic features and deep image features for recognition. Shahid et al. [24] obtained the candidate flame regions by combining the shape features and the motion stroboscopic features of the flame and then used the classifier to identify them. Zhang et al. [25] improved the target detection network YOLOv5 by combining the static and dynamic features of the flame, and solved the problem of unbalanced positive and negative samples. Yuan et al. [26] detected the forest wildfire by combining the deep static spatial features with deep dynamic features, and achieved good detection results. Wang et al. [27] proposed to integrate the bottom color features and motion features of the flame to design a multistage flame detection method, but failed to detect the flame object in real time. As shown above, most research on flame-based detection focuses on dynamic detection, which can achieve better results than static detection [28].

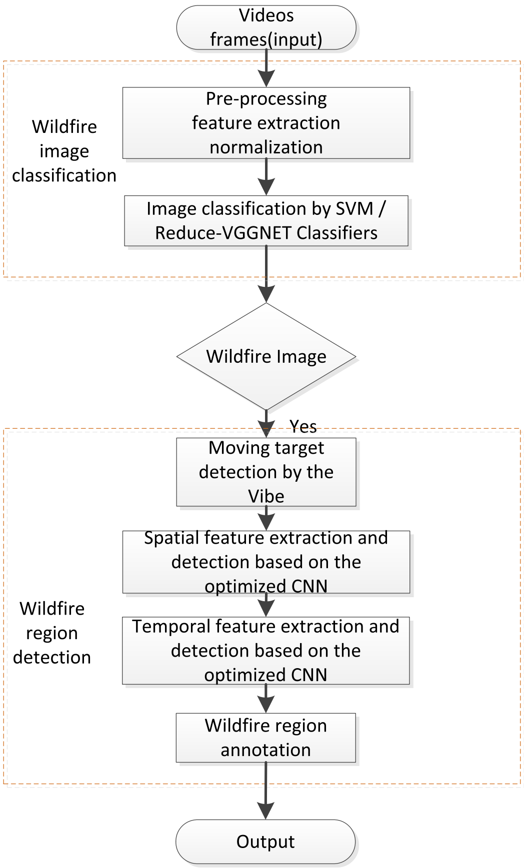


FIGURE 1. The framework of forest wildfire region detection.

Besides the above research, lost of studies use machine learning and deep learning algorithms to detect wildfire. Most of the recent detection algorithms use Convolutional Neural Networks (CNNs), including different versions of YOLO, U-Net, and DeepLab [33]. For example, Rashkovetsky et al. [34] proposed a single-input convolutional neural network based on the well-known U-Net architectures to detect wildfire area in satellite images. Sousa et al. [2] proposed a transfer learning-based method and data augmentation for wildfire early warning. Treneska and Stojkoska [36] also utilized transfer learning by fine-tuning the ResNet50 to detect wildfire on UAV collected images, which can obtain 88% accuracy. Jindal et al. [35] utilized an algorithm based on YOLOv3 and YOLOv4 to detect forest smoke. The results show that YOLOv3 outperforms YOLOv4 in all evaluation metrics. But, the above methods cannot obtain satisfactory detection results, which may be further improved by parameter optimization and different data augmentation techniques.

III. DESIGN OF THE EXPERIMENT

To enable students to master both traditional machine learning and deep learning and enhance the accuracy of wildfire detection, this paper divides the forest wildfire detection task into two modules, namely, wildfire image classification and wildfire region detection (see Figure 1). At the same time, we propose a novel wildfire image classification algorithm based on Reduce-VGGnet, and a wildfire detection algorithm based on the optimized CNN with the combination of spatial and temporal features. The wildfire image classification module extracts the video image frame, and then extracts the shape, texture and color features of the images and normalizes them. Then, we design a wildfire image classification algorithm based on traditional machine learning (SVM) and Reduce-VGGNet. Finally, the wildfire region detection module requires further annotating the fire regions on the classified wildfire images, and uses the Vibe algorithm to detect the candidate fire region, and needs to design an optimized CNN to extract temporal and spatial features respectively for wildfire region detection. This comprehensive experiment is designed hierarchically from different angles, which is consistent with the students' gradual understanding and cognitive process of image processing and image recognition.

A. WILDFIRE IMAGE CLASSIFICATION

The purpose of this module is to let students understand and master the image preprocessing, image feature extraction, and image classification. Before classification, we need to extract the image frames and extract features from the images. In this experiment, we propose the Reduce-VGGNet module to classify forest wildfire images, and require students to compare this method with the traditional machine learning algorithm SVM.

Preprocessing First, we extract the video image frames by the OpenCV module. OpenCV has powerful video editing capabilities, and encapsulates many image processing API functions, including image reading, scanning and face recognition. To extract meaningful information from a video or image, the CV module, VideoCapture(File_path), read() and imwrite(filename, img[, params]) functions can be used for video reading, and the image frame can be saved to a specified file.

Feature extraction Next, we extract image features based on color, texture and shape features. To extract the color features, we transform the RGB image to a gray image and extract the gray histogram features, including the mean and standard deviation of brightness and the probability of gray value. To extract the texture features, we extract the gray co-occurrence matrix, and extract seven invariant moments based on the co-occurrence matrix. For the shape features, the area, roundness, boundary circumference and boundary roughness of fire region are extracted. The above feature extraction belongs to the digital image processing part of Machine Vision. Students can further deepen their understanding of the basic knowledge of digital image processing through this module.

Normalization Due to the different ranges of different features, it is necessary to normalize them to the range of [0,1] to accelerate the convergence speed of the algorithm.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where x' denotes the data after normalization, x denotes the extracted feature, x_{min} denotes the minimum value of one feature, and x_{max} denotes the maximum value of one feature.

Finally, we input the normalized features to SVM, and compare the performance of SVM and Reduce-VGGNet.

1) SVM

SVM is a very popular classifier in recent years. It can realize nonlinear segmentation of feature vectors. Kernel functions in SVM can simplify the number of inner product calculations, reduce the running time, and can convert the inner product of high-dimensional space to a low-dimensional space. The performance of support vector machine mainly depends on the selection of kernel function, and the selection of kernel function depends on the actual dataset. In addition, students are required to determine the penalty factor c that affects the generalization ability of the classifier and the parameters of the kernel function g through cross-validation. The parameters with the best performance on the training set are selected as the final parameters of the model.

The Kernel functions in our experiment mainly include Radial Basis Function (RBF), Polynomial Kernel and Sigmoid kernel. We use the package Libsvm to establish the classification model. The process of SVM classification is shown in the figure below:

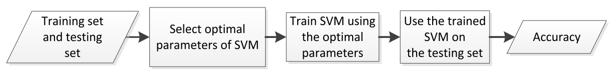


FIGURE 2. The flow chart of the SVM classification.

2) REDUCE-VGGNET

The Reduce-VGGNet model proposed in this experiment takes VGG-16 as the basic network structure. VGG16 network consists of 13 convolutional layers and 3 full connected layers. After each group of convolutional layers, a max pooling layer is connected, followed by ReLU activation function to solve the gradient dispersion problem.

As a deep convolution neural network, VGGNet has been widely used in image classification tasks. Based on the idea of transfer learning, this experiment transfers the weight parameters obtained from the training set of the network to the wildfire image set. As shown in Figure 3, the weight coefficients of the first 13 layers are transferred, the original three full connected layers are removed, and two full connected layers and softmax are used instead. The number of neurons of two full connected layers are set to 1024 and 2 respectively. We use the forest wildfire image dataset to train the full connected layers and Softmax classifier, and fine tune the VGG16 for classification. The purpose of this design is to reduce the training parameters and training time in VGGNet model.

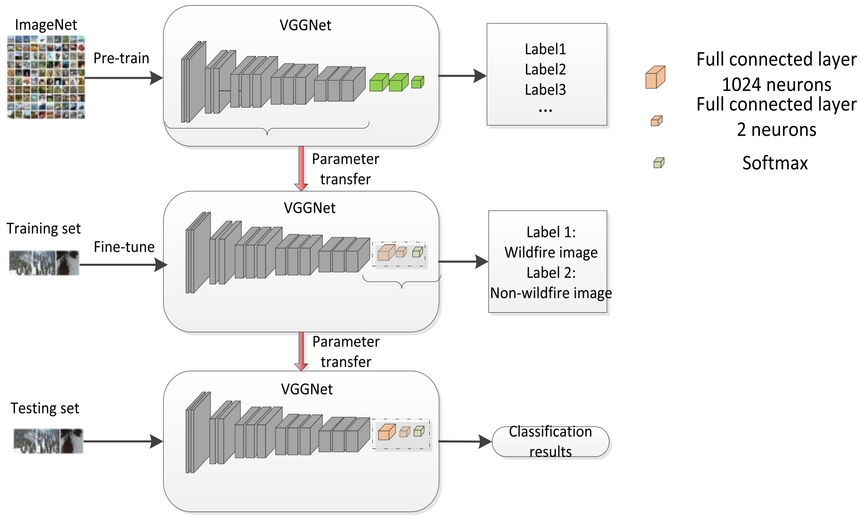


FIGURE 3. The structure of Reduce-VGGNet.

In this experiment, stochastic gradient descent (SGD) and Momentum are combined to train the model. We set the epoch to 100, the batch_size to 64, and the momentum parameter to $\beta_1 = 0.9$, $\beta_2 = 0.999$, the initial value of the learning rate is set to 0.001. We use cross entropy loss function to train the model:

$$Logloss = -\frac{1}{T}\sum_{t=1}^{T}(y_t \log(\hat{y}_t) + (1 - y_t)\log(1 - \hat{y}_t)) \quad (2)$$

where T represents the training sample, y_t is the expected category and $\hat{y}_t$ is the predicted category.

The learning rate of this experiment is updated by the exponential decay method:

$$\eta = \eta_0 \cdot \alpha^{(\lfloor l/d \rfloor)} \quad (3)$$

where η represents the updated learning rate, η_0 is the initial learning rate, α is the decay coefficient, and $\lfloor l/d \rfloor$ denotes the downward rounding of the quotient of the number of iterations and the decay step size. In the training process, the loss value is calculated after each learning rate is set. When the loss value is stable, the learning rate can be reduced, so that the minimum learning rate can be obtained by repeated experiments.

B. WILDFIRE REGION DETECTION

This module is designed to enable students to understand and master how to build a deep neural network model to extract spatial and temporal features. Thus, we can use these features to detect wildfire regions.

The wildfire region detection contains two stages, the first stage can detect moving objects by Vibe algorithm, the second stage uses 16*16 blocks to traverse the moving object and classify each of these blocks to be wildfire blocks or non-wildfire blocks. The detection of wildfire regions can be considered as a binary classification problem in our work. Considering that deep CNN architectures are the best choice for classification due to their ability of extracting high representative features, and also, the forest wildfires are occurring in different spatial and temporal scales, we design new CNNs to extract both spatial and temporal features in this paper.

The specific steps of this module are designed as follows: we set the moving objects detected by the Vibe algorithm as candidate wildfire regions, then we use 16 * 16 blocks to traverse these regions, and optimize the CNN network through appropriate network depth and multiple convolution kernel sizes to extract the spatial features of each block and classify them. If it is classified as a flame block, the region will be annotated. Furthermore, we extract the temporal features of candidate flame blocks by an optimized CNN model to annotate the wildfire region. We finally detect the wildfire region through the combination of temporal and spatial features.

1) VIBE ALGORITHM

Vibe is an effective object detection algorithm proposed by Barnich and Van Droogenbroeck [29]. First, we initialize the background model. Vibe algorithm needs to initialize the background model with the first frame of the video. For each pixel, considering that its adjacent pixels may have similar pixel values, the pixel value of its neighborhood is randomly selected as its sample value. Then, the background modeling and foreground detection are carried out. The main idea of this algorithm is to determine whether a pixel is the background point. Background modeling stores a sample set $M(x) = \{v_1, v_2, \ldots, v_n\}$ for each background point x, n is the size of the sample set. The pixel value of its neighbor point is randomly selected as its sample value. For each new pixel, we calculate the distance between the pixel and each value in the sample set. When this number is greater than the threshold T, the new pixel is considered as the background, otherwise it is considered as the foreground:

$$\#\{d(N_R(x), \{v_1, v_2, \ldots, v_n\})\} \geq T \quad (4)$$

TABLE 1. The structure of spatial CNN network.

Layer number	Type	Size	Step length	Input	Output
1	b			28*28*3	28*28*3
2	c	5*5	1	28*28*3	24*24*10
3	r			24*24*10	24*24*10
4	p	2*2	2	24*24*10	12*12*10
5	b			12*12*10	12*12*10
6	c	5*5	1	12*12*10	8*8*50
7	r			8*8*50	8*8*50
8	p	2*2	2	8*8*50	4*4*50
9	b			4*4*50	4*4*50
10	c	4*4	1	4*4*50	1*1*100
11	r			1*1*100	1*1*100
12	c	1*1	1	1*1*100	1*1*2
13	d				
14	s			1*1*2	1*1*2

where $N_R(x)$ represents the neighborhood with the radius R of the point x, # represents the number. Students can try different parameters in their experiments. For example, the parameter n can be set to 10, 20, and 30, the threshold T can be set to [1] and [10], and the threshold R can be set to [10] and [30], from which the optimal threshold can be selected. Finally, we update the background. The background model needs to be updated, and a value in the pixel sample set is randomly replaced with a new pixel value. In addition, we update the background according to a certain probability. When a pixel is determined as the background, we update it with the probability of $1/r$, where r is the time sampling factor, and we set it to 16 in our experiment.

2) SPATIAL FEATURE EXTRACTION

Firstly, we use 16 * 16 blocks to traverse the foreground area detected by the Vibe algorithm. Then, when the number of foreground pixels of the current block is greater than a certain threshold, an optimized CNN is designed to extract the features of the current block for classification.

As shown in the following figure, the upper left corner box in the left figure defines the region to be detected, and the right figure shows the moving foreground detected by the Vibe algorithm for this region. The small box in the upper left corner of the right figure represents a 16 * 16 block. We use the block to traverse from top to bottom for CNN forest wildfire detection. Before inputting the image blocks to CNN network, we normalize the 16 * 16 block size to 28 * 28 by the Bilinear Interpolation algorithm.

Convolutional Neural Network (CNN) is a commonly used deep learning architecture with convolution, which is an important model for image recognition speech recognition [37] and other issues [38], [39]. Compared with general neural networks, its three main characteristics are local receptive field, weight sharing and pooling. The design of CNN convolution neural network is the difficulty of this experiment, and it is necessary to design an optimization model that adapts to the characteristics of forest wildfire images. To extract spatial features of image blocks, we design the CNN shown as follows:

FIGURE 4. An example of candidate flame area traversal.

As shown in Table 1, in the "Type" column of Table 1, "b" represents the batchnormalize layer, "c" represents the convolution layer, "r" represents the Relu layer, "p" represents the pooling layer, "d" represents the dropout layer, and "s" represents the softmax layer. The batchnormalize layer is set before the convolution layer to normalize the data in batches. Therefore, layers 1, 5, and 9 are all set as the batchnormalize layer, followed by the convolution layer. Layers 2, 6, 10 and 12 are set as convolution layers. Layers 2 and 6 use a 5*5 convolution kernel, layer 10 uses a 4*4 convolution kernel, and layer 12 uses a 1*1 convolution kernel. It is used to map 100 featuremaps to 2-dimensional vectors for binary classification. Layers 3, 7 and 11 are designed as the ReLU activation layer and placed after the convolution layer. Layers 4 and 8 are set as the max pooling pool layer to expand the receptive field. Layer 13 is set as the dropout layer, and the drop factor is set to 0.5 in this experiment to enhance the generalization ability of the model. Layer 14 is

TABLE 2. The structure of temporal CNN network.

Layer	Type	Size	Step length	Input	Output
1	b			28*28*10	28*28*10
2	c	5*5	1	28*28*10	24*24*20
3	r			24*24*20	24*24*20
4	p	2*2	2	24*24*20	12*12*20
5	b			12*12*20	12*12*20
6	c	5*5	1	12*12*20	8*8*50

TABLE 3. The detailed information of our dataset.

Type	Training set	Validation set	Testing set	Total
Wildfire images	540	180	180	900
Non-wildfire images	600	200	200	1000
Total	1140	380	380	1900

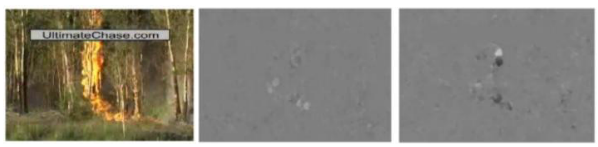
the softmax layer. The "size" column in the table means the size of convolutional kernel. The "step" column means the stride step. The "input" column means the shape of the input tensor, while the "output" column means the shape of the output tensor.

3) TEMPORAL FEATURE EXTRACTION

This module intends to extract optical flow sequence features as the input of CNN model to extract dynamic temporal features. The optical flow represents the instantaneous speed of a moving object. The optical flow is defined as the displacement vector field $\boldsymbol{d}_t$ of anterior and posterior frames, where $\boldsymbol{d}_t(m, n)$ represents the displacement vector of the pixel (m, n) from time t to time $t+1$, $\boldsymbol{d}_t^x$ represents the horizontal component and $\boldsymbol{d}_t^y$ represents the vertical component. In order to show its temporal feature, for a group of continuous L-frame optical flows, the horizontal component $\boldsymbol{d}_t^x$ and the vertical component $\boldsymbol{d}_t^y$ are concatenated as a 2L-channel optical flow sequence, which is denoted as $\boldsymbol{d}_t^{x,y} \in \mathbb{R}^{w*h*2L}$.

Figure 5 shows a wildfire image from two continuous images and its corresponding optical flow fields. Figure (b) represents the optical flow field in the horizontal direction, and Figure (c) represents the optical flow field in the vertical direction. It can be seen that the optical flow field in the flame region varies greatly, but the optical flow field in other regions with little change is relatively smooth. The flame movement is generally reflected in the vertical direction, so the change of optical flow field in the vertical direction is more significant than that in the horizontal direction.

We input $\boldsymbol{d}_t^{x,y} \in \mathbb{R}^{w*h*2L}$ to an optimized CNN network to extract the temporal features and classify them. The value of L is set as 5 in our experiment. The CNN network structure designed is similar to that in Table 1. As shown in Table 2, only the structures of layer 1 to 6 are different, and the parameters in the other layers are the same.



(a) (b) (c)

FIGURE 5. An example of the optical flow field.

IV. RESULTS AND ANALYSIS

For the experiment of wildfire image classification, the accuracy is used to evaluate our model [26].

$$\text{accuracy} = \frac{N_{pos}}{N_{total}} \tag{5}$$

where N_{pos} represents the number of images that are classified correctly, N_{total} represents the total number of images. In case of wildfire region detection, we use precision, recall and accuracy to evaluate. The calculation of accuracy is the same as Eq. (5), which can also be denoted as:

$$\text{accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{6}$$

The precision and recall can be denoted as follows:

$$\text{precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{7}$$

$$\text{recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{8}$$

where TP is True Positive, TN is True Negative, FP is False Positive, FN is False Negative.

In order to verify our experiment, we used the dataset from FLAME, a forest wildfire dataset opened by Northern

FIGURE 6. Examples of wildfire images.

FIGURE 7. Examples of non-wildfire images.

TABLE 4. The effects of normalization.

Normalization	Accuracy
Non-normalization	80.25%
Normalization	**83.64%**

TABLE 5. The effects of kernel functions.

Kernel	Accuracy
Radial basis kernel	**83.64%**
Polynomial kernel	82.28%
Sigmoid	81.82%

Arizona University [30]. A total of 900 consecutive images of wildfire flames were collected as positive samples and 1000 non-flame images were collected as negative samples from FLAME. We divided the dataset into three parts, which are training set, validation set and testing set. The wildfire region detection based on the optimized CNN needs to identify the specific location of the wildfire, so it is necessary to manually annotate the fire region containing the wildfire. Each wildfire image was divided into 16*16 areas, we manually annotated each block in each image, while the blocks from the non-wildfire images can directly be annotated as non-wildfire blocks. In addition, we set the batch size of CNN to 100. For the wildfire classification experiment based on SVM and Reduce-VGGNet, it only needs to know whether the image is a positive sample or a negative sample. The detailed information about our dataset is shown in Table 3. Figure 6 shows some wildfire images. Figure 7 shows some non-wildfire images.

A. RESULTS OF WILDFIRE IMAGE CLASSIFICATION

Considering the requirements of VGG16 model on the input images, all images are scaled to 224×224 in our experiment. For wildfire image classification, we only use the training set and the testing set in Table 3.

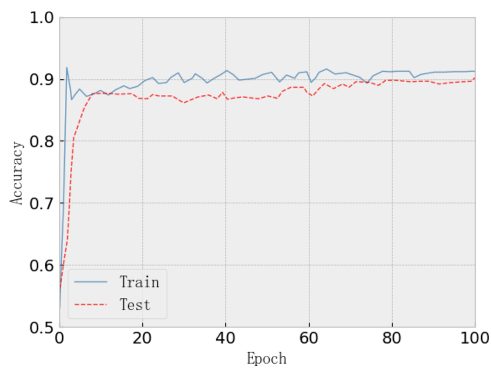


FIGURE 8. The accuracy curves of Reduce-VGGNet with different epochs.

1) RESULTS OF SVM

The parameters in libsvm are set to default values. The model is trained on 1140 images, and is tested on 380 images. The results of classification are closely related to the selection of parameters. First, we conduct experiments on the operation of non-normalization, and the results are shown in Table 4. It can be seen that the classification results after normalization are much higher than that after non-normalization. The accuracy is 83.64% after normalization.

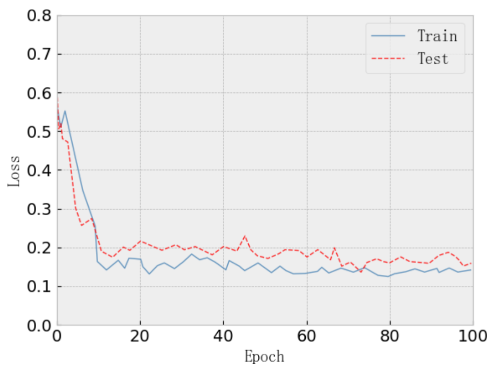


FIGURE 9. The loss curves of Reduce-VGGNet with different epochs.

Secondly, we compare the effects of different kernel functions after normalization, including radial basis kernel function, polynomial kernel function and Sigmoid kernel function. The results are shown in Table 5. It can be seen from Table 4 that the performance of Radial basis kernel is superior to Polynomial kernel and Sigmoid kernel.

Thirdly, for the selection of optimal parameters c, g, students are required to select the combination with the highest accuracy within the interval $[-10, 10]$, respectively with 0.5 step length for cross-verification, and select the combination c, g with the best performance. On the basis of this module, students can be encouraged to compare it with other algorithms, such as Adaboost.

2) RESULTS OF REDUCE-VGGNET

Figure 8 shows the experimental results of Reduce-VGGNet network. When the number of iterations is close to 100, the network reaches the convergence state, and the accuracy on the test set reaches 91.20%. It can be seen that the experimental result of this method is superior to that of SVM algorithm. Figure 9 shows the loss curves in model training and testing. The larger the number of iterations, the more the model tends to converge, and the convergence speed of this model is very fast.

B. RESULTS OF WILDFIRE REGION DETECTION

For the CNN-based spatial feature extraction and detection, after 1000 iterations on the validation set, the accuracy of wildfire region detection can reach 92.6%. For the CNN-based temporal feature detection, after 1000 iterations on the validation set, the accuracy of wildfire region detection can reach 93.92%. It can be seen from Table 6 that the accuracy of the model can reach about 88% after 100 iterations on both temporal and spatial features. As shown in Table 8, if the spatial and temporal features are combined, the accuracy can reach 97.35% under the classification operation, and can reach 94.37% without the classification operation. This result shows that the classification operation before wildfire region detection can help improve the performance. Figure 10 shows some examples of the detection of fire regions. It can be seen from the experimental results that the CNN optimization model can effectively detect the wildfire region.

TABLE 6. The results of spatial feature detection based on the optimized CNN.

Iteration times	Accuracy
100	88.25%
200	88.34%
300	89.64%
400	91.58%
500	91.12%
600	91.45%
700	91.89%
800	92.05%
900	92.52%
1000	**92.60%**

TABLE 7. The results of temporal feature detection based on the optimized CNN.

Iteration times	Accuracy
100	88.35%
200	88.69%
300	90.24%
400	91.62%
500	91.79%
600	91.32%
700	92.78%
800	93.95%
900	93.78%
1000	**93.92%**

TABLE 8. The results of the combination of spatial and temporal features.

Type	Accuracy
Detection with classification	**97.35%**
Detection without classification	94.37%

Moreover, we compare our method with well-known wildfire detection algorithms proposed by Habiboğlu [31] and Oh [32] on testing set. The former research divides the video into some blocks and adopts covariance-based features extracted from these blocks to detect fire regions. The latter research adopts a light weight CNN model to tackle the early detection wildfire region. These two methods have detailed steps, which are suitable for students to conduct experimental comparison. We also compare our method with some state-of-the-art object detection algorithms. We adopt Faster R-CNN (Faster Region-based Convolutional Network) and SSD (Single Shot Multibox Detector) models [40] for comparison. The results are shown in Table 9. It can be seen that the method proposed in our manuscript performs better than the other methods. The accuracy of our method is 2.34% higher

FIGURE 10. The results of some detected wildfire regions.

TABLE 9. The experimental results of different methods.

Method	precision	recall	accuracy
Our method	97.22%	97.22%	97.35%
Habiboğlu	92.78%	93.29%	93.85%
Faster R-CNN	94.44%	94.97%	95.01%
SSD	83.33%	79.36%	81.84%
Oh	88.33%	86.88%	88.20%

than that of Faster R-CNN, and 3.50% higher than that of Habiboğlu.

V. CONCLUSION

This paper designs a comprehensive experiment of forest wildfire detection, which covers digital image processing, machine learning and deep learning, and meets the requirements of the comprehensive experiment for Machine Vision course. The topic selection of this experiment is still a hot topic in current image processing research. The designed experiments focus on wildfire image classification and wildfire region detection. We propose the Reduce-VGGNet model for classification, and propose the combination of spatial and temporal features based on the optimized CNN model for wildfire region detection. The experimental results show that the Reduce-VGGNet can achieve better results than SVM, the optimized CNN under classification operation can achieve better performance than other state-of-the-art methods. Students can further explore the recent research on the basis of this experiment, and further improve our methods to achieve more innovative improvement in the field of forest wildfire detection. The experiment effectively combines teachers' scientific research with teaching. It can not only mobilize students' subjective initiative, but also improve their practical ability and innovative thinking, to meet the needs of fostering talents in machine vision for artificial intelligence.

Our framework may not be directly employed to satellite images. Although the combination of spatial and temporal features in our method can solve the problem of wildfires occurring in different spatial and temporal scales, there are still some challenges in satellite data processing, such as the complexity of noise information (clouds, lighting, smoke, etc.). Traditional deep learning frameworks may not solve these challenges.

In our future work, 1) We will explore other CNN-based methods and some pre-processing methods to further improve both the speed and the accuracy of wildfire region detection. 2) We will utilize satellite images and explore the detection of wildfire at pixel level. We can explore multi-sensor data from different sources (openly available satellite imagery) to improve the results in noisy conditions. Specifically, we can combine the single-sensor detection results and quantify its improvement.

REFERENCES

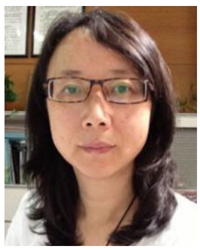
[1] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: A review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 927–948, 2019.
[2] M. J. Sousa, A. Moutinho, and M. Almeida, "Wildfire detection using transfer learning on augmented datasets," *Exp. Syst. Appl.*, vol. 142, Mar. 2020, Art. no. 112975.
[3] L. Wang, Y. Zhang, and X. Xu, "A novel group detection method for finding related Chinese herbs," *J. Inf. Sci. Eng.*, vol. 31, no. 4, pp. 1387–1411, 2015.
[4] B. Tang, J. Kong, and S. Wu, "Review of surface defect detection based on machine vision," *J. Image Graph.*, vol. 22, no. 12, pp. 1640–1663, 2017.
[5] S. Yin, Y. J. Ren, T. Liu, S. Guo, J. Zhao, and J. Zhu, "Review on application of machine vision in modern automobile manufacturing," *Acta Optica Sinica*, vol. 38, no. 8, 2018, Art. no. 0815001.
[6] F. Min and T. Lu, "Exploration and practice of machine vision course teaching for production practice," *Comput. Educ.*, vol. 10, pp. 41–43, Jan. 2017.
[7] Z. Wang, G. Xiao, and H. Liu, "Teaching reform of the course of machine vision principles and applications for professional degree postgraduates," *Educ. Teaching Forum*, vol. 3, no. 11, pp. 37–40, 2021.
[8] T. Shao, J. Tang, and Y. Liao, "Comprehensive experiment system of COCO on sorting based on machine vision," *Experim. Technol. Manag.*, vol. 37, no. 289, pp. 125–127&132, 2020.
[9] Z. Han and X. Liu, "Design of machine vision experiment teaching platform based on Python," *Comput. Meas. Control*, vol. 258, no. 3, pp. 249–253, 2020.
[10] J. Sigut, M. Castro, R. Arnay, and M. Sigut, "OpenCV basics: A mobile application to support the teaching of computer vision concepts," *IEEE Trans. Educ.*, vol. 63, no. 4, pp. 328–335, Nov. 2020.
[11] M. Cote and A. B. Albu, "Teaching computer vision and its societal effects: A look at privacy and security issues from the students' perspective," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1378–1386.
[12] S. Spurlock and S. Duvall, "Making computer vision accessible for undergraduates," *J. Comput. Sci. Colleges*, vol. 33, no. 2, pp. 215–221, 2017.
[13] Y. Jiang and B. Li, "Exploration on the teaching reform measure for machine learning course system of artificial intelligence specialty," *Sci. Program.*, vol. 2021, Nov. 2021, Art. no. 8971588.
[14] Y. Wang, M. Stankovic, A. Smith, and E. T. Matson, "Leader-follower system in convoys: An experimental design focusing on computer vision," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Aug. 2021, pp. 1–6.
[15] N. J. Abram, B. J. Henley, A. Sen Gupta, T. J. R. Lippmann, H. Clarke, A. J. Dowdy, J. J. Sharples, R. H. Nolan, T. Zhang, M. J. Wooster, J. B. Wurtzel, K. J. Meissner, A. J. Pitman, A. M. Ukkola, B. P. Murphy, N. J. Tapper, and M. M. Boer, "Connections of climate change and variability to large and extreme forest fires in Southeast Australia," *Commun. Earth Environ.*, vol. 2, no. 1, p. 8, Jan. 2021.

[16] O. M. Bushnaq, A. Chaaban, and T. Y. Al-Naffouri, "The role of UAV-IoT networks in future wildfire detection," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16984–16999, Dec. 2021.
[17] G. Saldamli, S. Deshpande, K. Jawalekar, P. Gholap, L. Tawalbeh, and L. Ertaul, "Wildfire detection using wireless mesh network," in *Proc. 4th Int. Conf. Fog Mobile Edge Comput. (FMEC)*, Jun. 2019, pp. 229–234.
[18] N. T. Toan, P. Thanh Cong, N. Q. V. Hung, and J. Jo, "A deep learning approach for early wildfire detection from hyperspectral satellite images," in *Proc. 7th Int. Conf. Robot Intell. Technol. Appl. (RiTA)*, Nov. 2019, pp. 38–45.
[19] L. Jia and Y. Jiong, "Multi-information fusion flame detection based on Ohta color space," *J. East China Univ. Sci. Technol.*, vol. 45, no. 6, pp. 962–969, 2019.
[20] Y. Tan, L. Xie, H. Feng, L. Peng, and Z. Zhang, "Flame detection algorithm based on image processing technology," *Laser Optoelectron. Prog.*, vol. 56, no. 16, 2019, Art. no. 161012.
[21] F. M. A. Hossain, Y. Zhang, C. Yuan, and C.-Y. Su, "Wildfire flame and smoke detection using static image features and artificial neural network," in *Proc. 1st Int. Conf. Ind. Artif. Intell. (IAI)*, Jul. 2019, pp. 1–6.
[22] T. Schultze, T. Kempka, and I. Willms, "Audio-video fire-detection of open fires," *Fire Saf. J.*, vol. 41, no. 4, pp. 311–314, Jun. 2006.
[23] Y. Xie, J. Zhu, Y. Cao, Y. Zhang, D. Feng, Y. Zhang, and M. Chen, "Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features," *IEEE Access*, vol. 8, pp. 81904–81917, 2020.
[24] M. Shahid, I.-F. Chien, W. Sarapugdi, L. Miao, and K.-L. Hua, "Deep spatial–temporal networks for flame detection," *Multimedia Tools Appl.*, vol. 80, nos. 28–29, pp. 35297–35318, Nov. 2021.
[25] D. Zhang, H. Xiao, J. Wen, and Y. Xu, "Real-time fire detection method with multi-feature fusion on YOLOv5," *Pattern Recognit. Artif. Intell.*, vol. 35, no. 6, pp. 548–561, 2022.
[26] J. Yuan, L. Wang, P. Wu, C. Gao, and L. Sun, "Detection of wildfires along transmission lines using deep time and space features," *Pattern Recognit. Image Anal.*, vol. 28, no. 4, pp. 805–812, Oct. 2018.
[27] Z. Wang, D. Wei, and X. Hu, "Research on two stage flame detection algorithm based on fire feature and machine learning," in *Proc. Int. Conf. Robot., Intell. Control Artif. Intell.*, Sep. 2019, pp. 574–578.
[28] J. Ryu and D. Kwak, "Flame detection using appearance-based pre-processing and convolutional neural network," *Appl. Sci.*, vol. 11, no. 11, p. 5138, May 2021.
[29] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.
[30] 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," *Comput. Netw.*, vol. 193, Jul. 2021, Art. no. 108001.
[31] Y. H. Habiboğlu, O. Günay, and A. E. Çetin, "Covariance matrix-based fire and flame detection method in video," *Mach. Vis. Appl.*, vol. 23, no. 6, pp. 1103–1113, 2012.
[32] S. H. Oh, S. W. Ghyme, S. K. Jung, and G. W. Kim, "Early wildfire detection using convolutional neural network," in *Proc. Int. Workshop Frontiers Comput. Vis.* Singapore: Springer, 2020, pp. 18–30.
[33] A. Bouguettaya, H. Zarzour, A. M. Taberkit, and A. Kechida, "A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms," *Signal Process.*, vol. 190, Jan. 2022, Art. no. 108309.
[34] D. Rashkovetsky, F. Mauracher, M. Langer, and M. Schmitt, "Wildfire detection from multisensor satellite imagery using deep semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7001–7016, 2021.
[35] P. Jindal, H. Gupta, N. Pachauri, V. Sharma, and O. P. Verma, "Real-time wildfire detection via image-based deep learning algorithm," in *Soft Computing: Theories and Applications* (Advances in Intelligent Systems and Computing), vol. 2, Singapore: Springer, 2021, pp. 539–550.
[36] S. Treneska and B. R. Stojkoska, "Wildfire detection from UAV collected images using transfer learning," in *Proc. 18th Int. Conf. Informat. Inf. Technol.*, Skopje, North Macedonia, 2021, pp. 6–7.
[37] A. A. Abdelhamid, E.-S.-M. El-Kenawy, B. Alotaibi, G. M. Amer, M. Y. Abdelkader, A. Ibrahim, and M. M. Eid, "Robust speech emotion recognition using CNN+LSTM based on stochastic fractal search optimization algorithm," *IEEE Access*, vol. 10, pp. 49265–49284, 2022.
[38] N. Lopac, F. Hrzic, I. P. Vuksanovic, and J. Lerga, "Detection of non-stationary GW signals in high noise from Cohen's class of time-frequency representations using deep learning," *IEEE Access*, vol. 10, pp. 2408–2428, 2022.
[39] L. Wang, Y. Zhang, and K. Hu, "FEUI: Fusion embedding for user identification across social networks," *Int. J. Speech Technol.*, vol. 52, no. 7, pp. 8209–8225, May 2022.
[40] 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.* Cham, Switzerland: Springer, Sep. 2016, pp. 21–37.

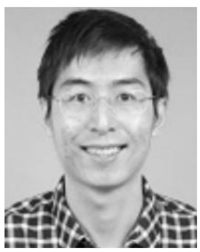
LIDONG WANG (Member, IEEE) was born in Wenzhou, Zhejiang, China, in 1982. She received the Ph.D. degree from the College of Computer Science and Technology, Zhejiang University, in 2013.

She was a Visiting Scholar with Tongji University, in 2022. She is currently an Associate Professor with Hangzhou Normal University. She has published more than 30 articles on social network analysis, data mining, pattern recognition, and computer education. Her current research interests include image processing, machine learning, and text mining.

HUIXI ZHANG received the master's degree in circuit and system from Zhejiang University, in 2005. She joined Hangzhou Normal University as a Lecturer. Her research interests include signal processing, system design, the Internet of Things technology.



YIN ZHANG was born in Lanzhou. She received the Ph.D. degree in computer science from Zhejiang University. She is currently an Assistant Professor with the College of Computer Science and Technology, Zhejiang University. Her current research interests include knowledge discovery, machine learning, digital library, and information and knowledge management.



KEYONG HU (Member, IEEE) received the Ph.D. degree in mechatronic engineering from the Zhejiang University of Technology, Hangzhou, China, in 2016. He is currently a Teacher of electronic information engineering with Qianjiang College, Hangzhou Normal University. His research interests include artificial intelligence and new energy technology.

KANG AN (Member, IEEE) received the master's degree in circuit and systems from the Guilin University of Electronic Technology, in 2007. He joined Hangzhou Normal University as an Associate Professor. His research interests include the Internet of Things technology and machine learning.