RESEARCH ARTICLE

PSFNet: A Deep Learning Network for Fake Passport Detection

JIAFENG XU[1], DAWEI JIA[1], ZHIZHE LIN[2], AND TENG ZHOU[3,4,5]

[1]Immigration Management College, China People's Police University (Guangzhou), Guangzhou 510663, China
[2]Office of Emergency Management, Shantou Central Hospital, Shantou 515041, China
[3]Department of Computer Science, Shantou University, Shantou 515063, China
[4]Key Laboratory of Intelligent Manufacturing Technology, Shantou University, Ministry of Education, Shantou 515063, China
[5]Guangdong Provincial Key Laboratory of Infectious Diseases and Molecular Immunopathology, Shantou 515063, China

Corresponding author: Zhizhe Lin (l.zhizhe@gmail.com)

ABSTRACT Passport security feature classification has always been an important issue in border checks. Current manual methods struggle to achieve satisfactory results in terms of consistency and stability for security features with high similarity. For this reason, this study designs and develops a deep learning-based passport security feature classification model that can identify similar security features. The proposed model is based on the ShuffleNet v2, and its network structure is further optimized by two aspects for our task. Firstly, we embed pixel attention in the model to enable the network to better focus on important features with discriminatory power. Secondly, we introduce focal loss to relieve the overfitting problem caused by data imbalance. Finally, the superiority of the proposed classification algorithm is verified with the constructed passport security features data. The classification accuracy of the proposed algorithm is enhanced by 0.8%. The experimental results show that the classification accuracy is as high as 95.5%.

INDEX TERMS Passport, security feature detection, deep learning.

I. INTRODUCTION

Fraudulent identity and travel documents can be linked to many criminal activities, including financial fraud, human trafficking, and terrorism [1]. Most of the current detection for passport authenticity relies on manual [2], which involves the devices such as UV lamp [3], IR detector *etc*. However, the efficiency of manual detection is not stable [4], which brings great challenges to border security. Therefore, designing a simple and fast passport security feature authentication method has important academic value and practical significance.

The development of artificial intelligence provides a new solution to this problem. Many scholars have designed automatic verification algorithms for security features based on deep learning. These algorithms are mainly used to verify the authenticity of identity documents [5], [6] and banknotes [7], which can improve the efficiency of detection. Although a few positive results have been achieved [8], [9], [10], there is still little research on algorithms used to detect travel documents.

The following three common issues remain to be addressed.

A. SIMILARITY OF PASSPORT SECURITY FEATURES

Figure 1 shows that the four categories of passport security features are similar in shape, with only subtle differences
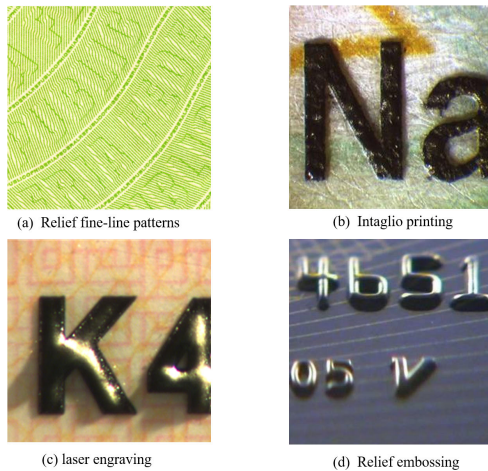
The associate editor coordinating the review of this manuscript and approving it for publication was Zhenhua Guo.

FIGURE 1. Four categories of passport security features were similar in shape, only with little difference between each other. Despite the fact that the number or the letter appears to be raised surface, they are all generated by separate procedures.

between each other. They are produced with different manufacturing techniques, despite the fact that the letter appears to be raised on surfaces. As one of the common anti-counterfeiting technologies, the relief fine-line pattern is designed according to the background pattern (logo or text) so that the pattern has an effect as if carved unevenly. The fine-line patterns on the background areas of the secure documents are difficult to reproduce without professional cutting-edge printing equipment [11]. Intaglio printing is a security feature that is adopted by a majority of the world countries to protect banknotes and other types of documents from being forged [12]. The patterns printed by intaglio printing are transferred from a metal surface to paper [13]. This technology is complex and costly. Intaglio printing is also often combined with other security features like blind signs and latent images. Laser engraving is the most effective technique for processing hard materials with complex geometries [14]. Pictures and text are engraved in plastic laminates or cards by means of a laser. Laser engraving does not need ink, it is a darkening process by burning special plastic with a laser ray. Most passport information pages are laser engraved with photos and personal information on multi-layer polycarbonate. During this process, a laser-sensitive film records the relevant data, allowing for the passport personalization process [15]. Relief embossing is also called blind embossing with colorless embossing of images or text, which involves the high-pressure embossing of letters, motifs, *etc*. It was a popular technique in the 17th and 18th centuries. Blind embossing is used for document protection in combination with polymer or foil nowadays. Tactile laminate features like fine line patterns or microprint are incorporated into security laminates by embossing [16]. As a result, how to choose the discriminating features for a deep convolution model in fine-grained images of document security technology has

become a pressing concern in document security technology image classification.

B. SAMPLE IMBALANCE

Nowadays, deep learning has been widely used in various fields, and many experiments have proven its reliability [17], [18], [19], [20]. Image features can be extracted by convolution pooling, and image recognition and classification can be performed on this basis [21]. However, these methods make the recognition results largely dependent on the quantity and quality of training samples [22]. Due to the limitation of capital and time, it is difficult to obtain a complete set of image datasets of security features on a large scale. However, class imbalance is still a big problem in passport detection. For example, there are a large number of relief fine-line patterns in a passport while the number of relief embossing is limited. Therefore, it has become a very challenging problem how to obtain high-accuracy grading of the passport security features under the condition of imbalanced samples.

C. LIGHTWEIGHT MODEL

Border guards must verify the authenticity of passports and infer whether there are any anomalies in the travel patterns of passengers [23]. This is critical for the early prevention of cross-border criminal activity. However, this analysis is often impossible due to a large number of passengers at the ports and the requirement for rapid processing of border procedures. Therefore, it is important to deploy a small training model on mobile terminals to enable faster and more accurate detection. Thus, it requires a lightweight training model with substantially fewer parameters for fast delivery and deployment.

In order to solve the above problems, we propose a Convolution Neural Network (CNN) model of the passport security features images. CNN can automatically collect image features of the passport security features. The improved model increases the classification accuracy of the passport features and further reduces the number of parameters and complexity while maintaining accuracy.

The contributions of this work are as follows:

- We propose a lightweight model called PFSNet for classifying passport security features. The improved model can realize higher accuracy with fewer parameters by embedding the pixel attention module to help enhance critical information and suppress useless features.
- We effectively relieve training overfitting by data augmentation and introducing focal loss functions to the model.
- We gather a particular passport security feature sample dataset comprising four groups:fine-line patterns, laser engraving, intaglio printing, and relief embossing. We compare five state-of-the-art methods on our dataset to evaluate the recognition accuracy. The experiment results demonstrate the superiority of the proposed method.

The remainder of this paper is organized as follows. We first discuss related works in Section 2. Section 3 describes the data preparation, image processing, model installation, and training of our experiments. In Section 4 the performance analysis is discussed. Finally, a conclusion with a few oncoming work purposes is explained in Section 5.

II. RELATED WORK

Research on document detection has focused on currency and passports. This section surveys the existing work and the most recent technologies on document security features categorization and image recognition and classification.

A. OPTICAL INSPECTION METHODS FOR PASSPORTS

In the field of document analysis, much research has been done to find ways to detect passports. Due to the widely used security materials in passports, many detections for security documents are based on optical methods. As a method of non-destructive testing, optical inspection has become an important way of passport detection. The security paper and security ink in the passport interact with the spectrum and excite different wavelengths of light. Abdelhameed et al. [24] developed a new security ink for security documents. The ink exhibits the best excitation wavelength at 360 nm, with a change in absorption color and fluorescence in the printed document. Purba et al. [25] use a variety of light sources of VSC8000HS to differ the Overlapped Black Pen Inks in Fraudulent Documents. Purba et al. [26] use the same equipment to differ different blue ink. Their study demonstrates the potential of multi-spectral illumination and micro-spectrometer for detecting handwritten documents. The acquired experimental findings show that the selected non-destructive testing methodology can be used to determine the authenticity of passports. Krol et al. [27] use Laser-induced breakdown spectroscopy (LIBS) to analyze Polish passports for forensic purposes. Their proposed method becomes an alternative to the traditional method. Marques et al. [28] analyze identification documents using optical coherence tomography imaging. They established its usefulness for quantitative visualization of embedded security features in these documents, increasing the accuracy of forgery detection. However, optical inspection relies on special instruments and requires border guards for manual detection, so it is not conducive to the fast deployment of the frontline.

B. IMAGE ANALYSIS BASED ON DEEP LEARNING

With the rapid development of deep learning, computer vision is widely used in many fields, such as agriculture, forensic science, and medicine. Deep learning can deal with a large number of complicated problems while maintaining higher speed, higher accuracy, and better robustness [29], [30]. Therefore, deep learning has great potential for image classification [31], target detection [32], and segmentation tasks [33]. However, there are still many challenges in image analysis. Numerous scholars have made many efforts to address these difficulties. Kim et al. [34] propose a new segmentation technique for weakly-supervised semantic segmentation, which the class-specific knowledge directly from the classification network by exploiting the image masking technique, thus solving the problem that the pseudo-label cannot accurately express the object regions and their classes. The recent work in the paper of Gao et al. [35] shows that a novel cloud-edge distributed framework has overcome the difficulty of transmitting massive data in the cloud-only deployment scheme, as well as the difficulty to analyze massive data in the edge-only deployment scheme. This method adopts a hierarchical information allocation strategy and a novel pyramidal deep learning model is proposed to effectively capture the global and local information of the salient object. In order to address the obstacles to Saliency detection, Gao et al. [36] introduce the co-saliency detection approach in Internet-of-Things (IoT). In their method, a multi-stage context perception scheme, two-path information propagation, and stage-wise refinement are adopted to address the inefficiency of inter-image information representation and single-image contextual information extraction.

C. PASSPORT DETECTION BASED ON DEEP LEARNING

Several studies have been performed using deep learning-based models for the classification and detection of passports. Jeny et al. [37] use ResNet50 to identify passport cover (98.56% accuracy). Zaaboub et al. [38] propose a framework that performs passport stamp detection and recognition with maximum accuracy of 94.5%, and this is the first classification method for passport stamps. Wang et al. [39] develop an improved algorithm based on the traditional SURF operator to achieve security thread detection in passports with an accuracy of 84.33%. Kim et al. [40] propose an effective recognition algorithm of the passport MRZ information using a combined neural network recognizer of Convolutional Neural Network and Artificial Neural Network. Liu et al. [41] present a specially designed model that is able to successfully extract MRZ information from digital images of passports of arbitrary orientation and size. The model achieves a 100% MRZ detection rate and a 99.25% character recognition macro-f1 score on a passport and visa dataset.

A review of the literature in this area reveals that no previous study has presented a design for a passport security feature classification system. Our research applies computer vision technology to the classification of passport security technology, which can alleviate the cost of manual work and waste of resources. And we focus on lightweight networks, which facilitate the deployment of network models to mobile devices, such as enforcement recorders, and can effectively assist law enforcement officers in their daily work.

III. METHODOLOGY

A. RESEARCH PROCESS

First, we collect passport security feature images through public sources and build a dataset. Then, we improve the ShuffleNet v2 model by introducing the focal loss function and the pixel attention (PA) module. Finally, following data training, we evaluate the performance of our model.

B. DATA COLLECTION AND PROCESSING

The inputs of the dataset for model training are image patches of passport security feature images from openly accessible sources, such as the *Public register of authentic travel and identity documents online* (PRADO) and the dataset of Pacific Security Technologies.

PRADO is a repository of passport information from the European Union, which collects documents from the EU and its member states, as well as from countries that have joined the PRADO program, and the online system is open to the public.

Some images are also selected from the dataset of Pacific Security Technologies. The dataset collects document samples from 212 countries/regions (2680 passports, 174 visas) and 199 countries/regions with currencies, and as many as 287,114 images. There are many partial detail images of passports in the sample library. The above sample library of these public channels collects a large number of passport security feature images, which provides rich data for our study.

The inclusion criteria in data collection are as follows:

1) The resolution of the picture should be greater than 200 * 200. 2) The light source for collecting pictures should be natural light or side light. 3) The object of the image should be Intaglio printing, relief fine-line pattern, Laser engraving, or Relief embossing.

We rotate the sample pictures at 90°, 180°, and 270°, flip them horizontally and vertically, and add random rectangles to avoid the over-fitting caused by the limited number of samples in the training process. As a result, we successfully expand the dataset and improve the generalization capacity of the model. Some images of the dataset are illustrated in Figure 2. Finally, this dataset contains 8148 images split into four classes.

The dataset after data enhancement is shown in Table 1. The labels are set to 0, 1, 2, and 3 for the dataset types. The number of laser engraving training sets with label 2 is 1302 after enhancement, while the number of Relief embossing is 1120 after enhancement, which is much smaller than the number of other labels, resulting in an unbalanced training set. The processed sample pictures were trained to construct a classification and recognition system of passport security features.

C. THE DETAILS OF THE PROPOSED MODEL

Compared with heavyweight networks, lightweight networks are characterized by fewer parameters, less computation, and
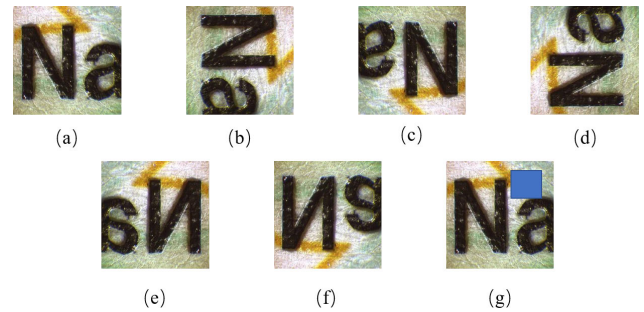

FIGURE 2. Seven samples from the dataset. (a) The original image, (b) The original image rotated at 90°, (c) The original image rotated at 180°, (d) The original image rotated at 270°, (e) The original image is flipped horizontally, (f) The original image is flipped vertically, (g) The original image is added random rectangle.

TABLE 1. Number of four kinds of samples in the training set.

Label	Label name	Before enhancement	After enhancement
0	Relief fine-line patterns	338	2366
1	Intaglio printing	480	3360
2	Laser engraving	186	1302
3	Relief embossing	160	1120

shorter inference time. It is more suitable for scenarios with limited storage space and power consumption, such as mobile embedded devices and other edge computing devices. Therefore lightweight networks have received wide attention, and representative networks include SqueezeNet, MobileNet v2, and ShuffleNet v2.

Although both MobileNet v2 and SqueezeNet have less number of parameters, the computation of SqueezeNet increases a lot, which affects the model computing time to some extent. On the other hand, MobileNet v2 presents linear bottlenecks and inverse residuals. Although it can obtain higher accuracy, the inference speed is slower than that of MobileNet v1.

The channel split proposed by ShuffleNet v2 makes the model more efficient. Each building block allows the use of a larger number of feature channels and a larger network capacity, while half of the feature channels in each building block pass directly through the block and join the next block. The amount of feature reuse decays in an exponential manner with the distance between two blocks, with feature reuse becoming weaker between distant blocks. The ShuffleNet v2 architecture implements this pattern of feature reuse by design, thereby increasing the efficiency of the model. The experimental results show that the accuracy of ShuffleNet v2 is higher than MobileNet v2.

The focus of this study is to apply the model to mobile devices, which requires considering both the recognition accuracy and speed of the model. By analyzing and comparing various lightweight networks, we concluded that ShuffleNet v2 is an efficient lightweight model with better recognition accuracy than other lightweight models at the same complexity. Therefore, ShuffleNet v2 is chosen
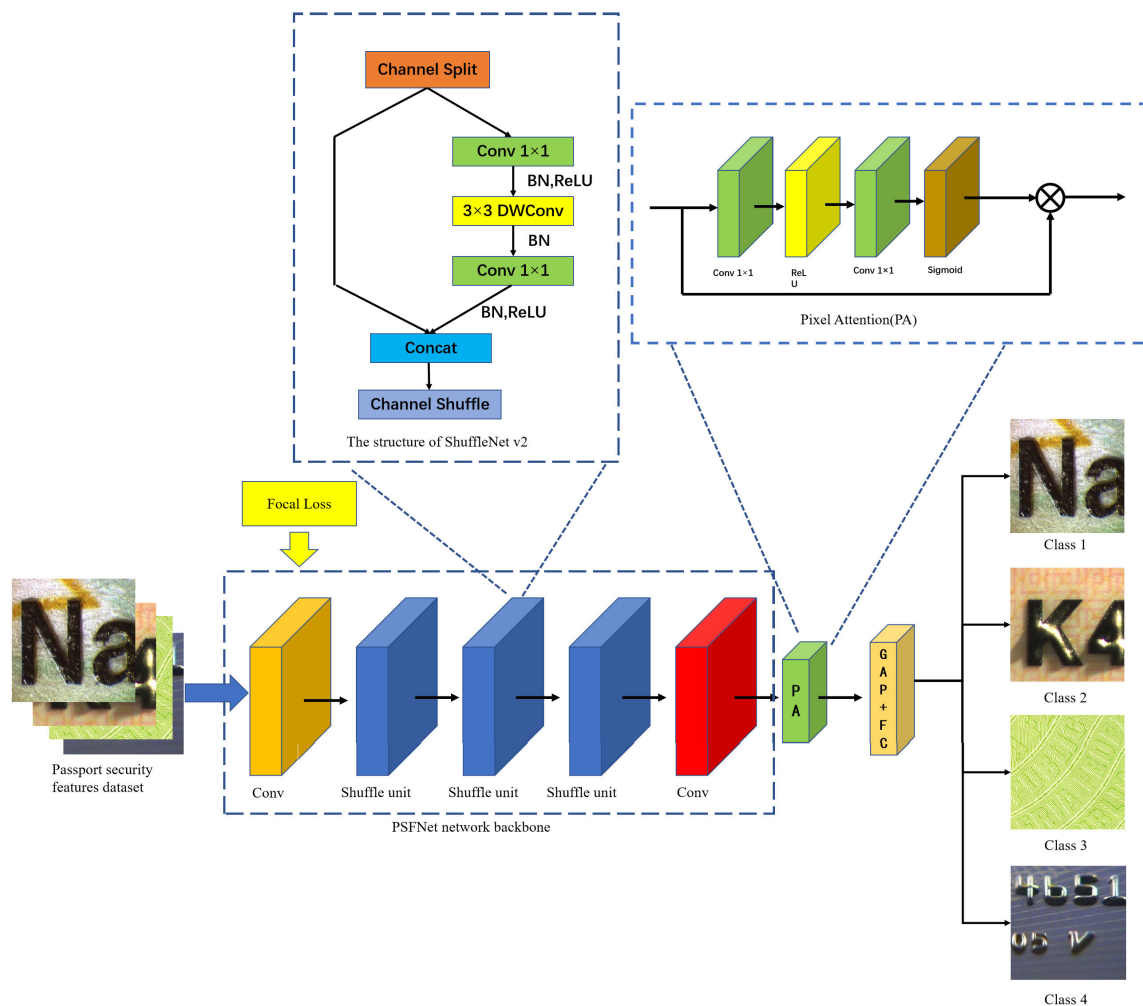
FIGURE 3. PFSNet is built on the basis of Shufflenet v2. The network structure is further optimized by Pixel Attention (PA) and Focal Loss. PA can improve the expressiveness of convolution and help generate better reconstruction results. On the other hand, Focal Loss can relieve the overfitting problem caused by data imbalance.

as the backbone network of the model. We make a series of improvements to the ShuffleNet v2 network to design a low-cost, high-accuracy end-to-end passport security feature classification model.

The architecture of the model is shown in Figure 3. We select ShuffleNet v2 as the core of the backbone feature extraction network. Then further embed a pixel attention mechanism in the model to enhance the ability of the network to fine-tune so that important features are enhanced while those less important are restrained. Furthermore, we introduce the focal loss function, which avoids low classification accuracy caused by the sample classification imbalance.

1) ShuffleNet v2

ShuffleNet v2 is proposed by Ma et al. [42] in 2018, which is an upgraded version network based on ShuffleNet v1. The structure of ShuffleNet v2 is shown in Figure 3. ShuffleNet v2 summarizes four design essentials of lightweight networks from practice, guided by practical inference speed,

and proposes ShuffleNet v2 according to the essentials with a balance between accuracy and speed, in which the channel split operation divides the input features into two parts and achieves the feature reuse effect similar to DenseNet.

At the beginning of each unit, the input of the c-feature channels is divided into two branches, one of which is left unchanged and the other one consisting of three convolutions that have the same input and output channels. The two 1×1 convolutions are no longer group convolutions but change to a normal 1×1 convolution operation. After convolution, the two branches are concatenated rather than added. Therefore, the number of channels remains the same. Then, the same "channels shuffle" operation as in ShuffleNet v1 is used to enable the communication of information between the two branches.

ShuffleNet v2 uses matrix stitching instead of *Add* operation in ShuffleNet v1, which not only retains the deep convolution to ensure certain accuracy but also retains the Channel Shuffle operation to exchange information between different

channels. The final experimental results also show that Shuf-fleNet v2 achieves a good trade-off between accuracy and efficiency in image classification and object recognition.

2) PIXEL ATTENTION MODULE

Although the design of pixel attention (PA) is inspired by that of channel attention (CA)and spatial attention (SA), it is experimentally shown that PA can improve the expressiveness of convolution and help generate better reconstruction results. In addition, PA contains fewer parameters, therefore PA is more suitable for lightweight models.

Channel attention module [43] first obtains a one-dimensional ($C \times 1 \times 1$) attention feature vector by global pooling, where C refers to the number of channels. In order to reduce the complexity of the model and improve the generalization ability, a bottleneck structure with two convolutional layers is used, where the first convolution layer plays the role of dimensionality reduction. The $1 \times 1 \times C$ vector first reduces the dimensionality of the input to $1/r$, then the ReLU activation function is applied, and The final convolutional layer brings the input back to the original dimension. Finally, the scale factor is normalized to the range [0, 1] by the sigmoid function. The whole process can be seen as learning the weight coefficients of each channel, thus making the model more discriminative of the features of each channel.

The structure of pixel attention [44] is quite similar to that of channel attention. However, different from channel attention and spatial attention, pixel attention is able to generate a 3D ($C \times H \times W$) matrix as the attention features. Pixel attention mechanism removes the global pooling layer [45], the rest of the structure is the same as channel attention. In PA, the input which has a dimension of $H \times W$ and contains C feature maps will change to $H \times W \times C/r$ and then rise back to $H \times W \times C$. Note that C refers to the number of channels, and H and W are the height and width of the features. Finally, the scale factor is normalized to the range [0, 1] by the sigmoid function. The PA module treats each pixel in the image differently, and it can learn the informative contextual feature of each pixel, allowing it to pay more attention to those pixels that are significant.

Through the above process, the pixel attention can learn the informative contextual feature of each pixel [46] and contains very few parameters but helps enhance the representational ability of the features. Therefore, we add PA after the feature extraction layer to improve the performance of ShuffleNet v2.

3) FOCAL LOSS

Since some security features are widely used in different national passports, while others are rarely used. The number of security features collected varies widely, leading to the problem of sample imbalance in the dataset. To solve this problem, we introduced focal loss.

The focal loss [47] was proposed in 2017 and was initially used in the image domain to solve model performance problems caused by data imbalance. Many evaluation results revealed that the focal loss function improved

the classification accuracy [48], [49], [50]. The focal loss is obtained based on a modification of the standard cross-entropy loss, which allows the model to focus more on the hard-to-classify samples during training by reducing the weights of the easy-to-classify samples. It adds weights to the losses corresponding to the samples according to the ease of sample discrimination. Add smaller weights α_1 to easily distinguishable samples while adding larger weights to hard-to-discriminate samples α_2. The expression of the loss function can be written as:

$$L_{sum} = \alpha_1 L_{easy} + \alpha_2 L_{hard} \qquad (1)$$

As α_1 is so small that the L_{hard} dominates the loss function. That is, the loss function is focused on the hard-to-discriminate samples. The loss function can be written as:

$$FL(p_t) = -(1 - p_t)^\gamma log(p_t), \qquad (2)$$

where $(1 - p_t)^\gamma$ is called modulating factor, which makes the model more focused on hard-to-classify samples during training by reducing the weights of easy-to-classify samples. When a sample is misclassified, p_t is small, then the modulation factor $1 - p_t$ is close to 1 and the loss is not affected. This function can be of great use in dealing with class imbalance problems. In the training of the deep model, the training effect of focal loss is better than that of the cross-entropy loss. Therefore, the loss function is replaced with focal Loss in the network.

IV. EXPERIMENTS

A. EVALUATION CRITERIA

In this study, we use the top 1 accuracy, precision, recall value, F1 score, and the number of network parameters as the evaluation criteria. The corresponding formula is described as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

$$precision = \frac{TP}{TP + FP} \qquad (4)$$

$$recall = \frac{TP}{TP + FN} \qquad (5)$$

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \qquad (6)$$

in which, TP is the true positive. FP is the false positive. TN is the true negative. FN is the false negative. The classification accuracy of a model is one of the important metrics to evaluate the overall performance of the model. However, in this task, the model needs to be deployed on a mobile device with limited random access memory (RAM). We use the number of parameters as another evaluation criterion.

B. EXPERIMENT PLATFORM

For all models, the parameters were optimized by the Adam optimizer with an initial learning rate of 0.001 and a mini-batch size of 128. We set the size of random clipping images to $224 \times 224 \times 3$, and train the model for
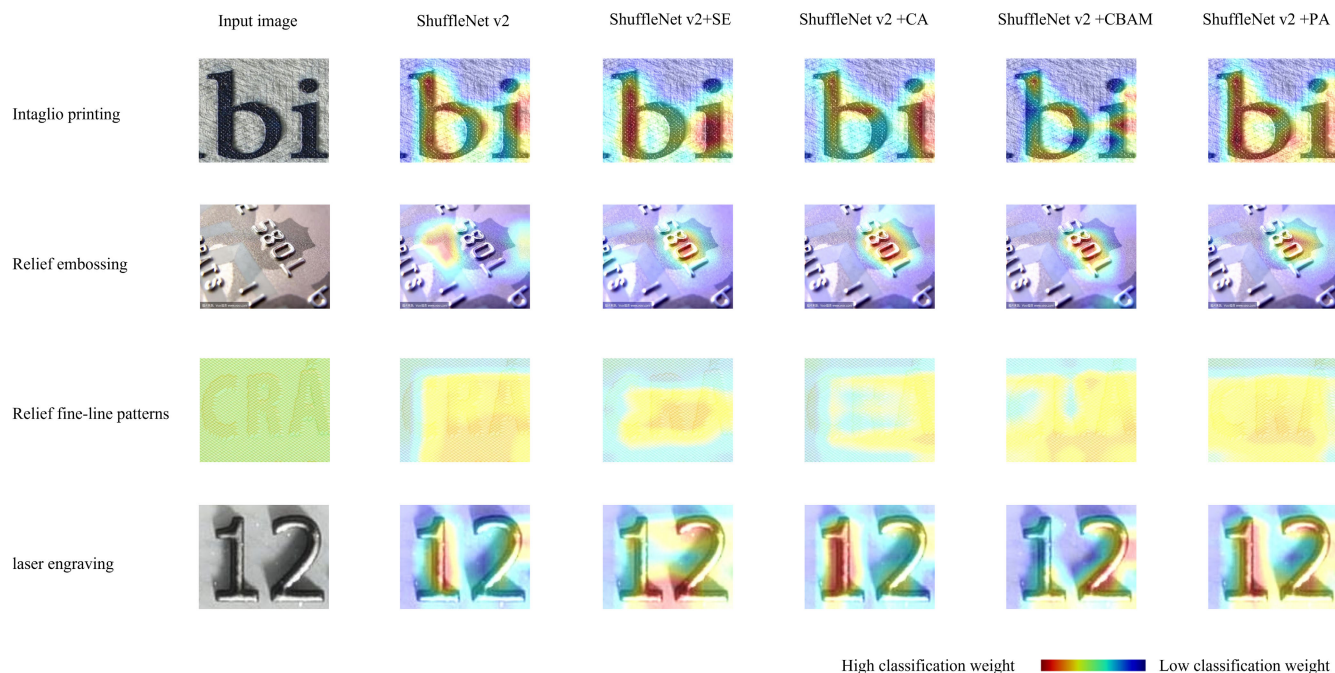
FIGURE 4. Grad-CAM maps of ShuffleNet with different attention blocks for four-classes passport security features image classification.

TABLE 2. Ablation study of different attention blocks embedded in the ShuffenNet v2.

Attention Embed	None	SE	CA	CBAM	PA
Accuracy (%)	94.7%	94.7%	94.8%	94.7%	94.9%

100 epochs. The source code is implemented by Python, including data preprocessing and algorithm implementation. All the models are conducted using Pytorch (version 1.7) and on a Ubuntu 18.04 LTS workstation with an Intel i7-8700K CPU@3.70GHz and a GPU of NVIDIA GTX10800Ti.

C. THE PERFORMANCE BY EMBEDDING DIFFERENT ATTENTION MODULES

In the ablation study, we embed four classical attention modules, *e.g.* Squeeze-and-Excitation Networks (SE), coordinate attention (CA), convolutional block attention module (CBAM), and pixel attention (PA) into ShuffleNet v2 respectively, to compare the effect of different attention mechanisms. The results of the classification accuracy show that embedding attention blocks into the model can effectively improve the classification accuracy of the passport identification task. Compared with SE, CA, and CBAM, PA achieves the highest accuracy. As shown in Table 2, the classification accuracy of ShuffleNet v2+*SE*, ShuffleNet v2+*CA*, and ShuffleNet v2+*CBAM* were 94.7%, 94.8%, and 94.7%, respectively, whereas the ShuffleNet v2+*PA* achieved the accuracy of 94.9%, 0.2% higher than the baseline model without the attention module.

Since PA can adaptively adjust pixel-level features in all channels, the model embedded with PA is able to focus more on important features in the classification task. Therefore, the above experimental results show that embedding attention blocks into the model can effectively improve the accuracy of the passport image identification task.

In this experiment, we introduce the Grad-CAM model [51] to better observe the distribution of classification weights of the model with embedded attention blocks for passport security feature image classification. Grad-CAM shows which parts of the image the model focuses on, and it uses the gradient information of the last convolution layer of the CNN model to assign importance to each neuron for specific attention decisions [52].

Taking the model embedded with different attention blocks as an example, the Grad-CAM plots of four different passport security feature images are shown in Figure 4.

The visualization results show that the distribution of classification weights of ShuffleNet v2 embedding different attention blocks is rather varied, suggesting that the attention blocks will affect the feature extraction process of the CNN model.

Firstly, the attention mechanism can help the model find the exact region for feature extraction. It can be observed that the distribution of the original ShuffleNet v2 classification weights in the relief embossing does not cover the text surface accurately. And after embedding the attention blocks in the model, the regions of classification weights are all able to accurately cover the surface of the security feature text.

No.	PA Block	Focal Loss	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
1			94.7%	94.1%	92.2%	92.7%
2	✓		94.9%	93.8%	92.5%	93.0%
3		✓	93.8%	92.8%	90.4%	90.6%
4	✓	✓	95.5%	94.5%	92.8%	93.1%
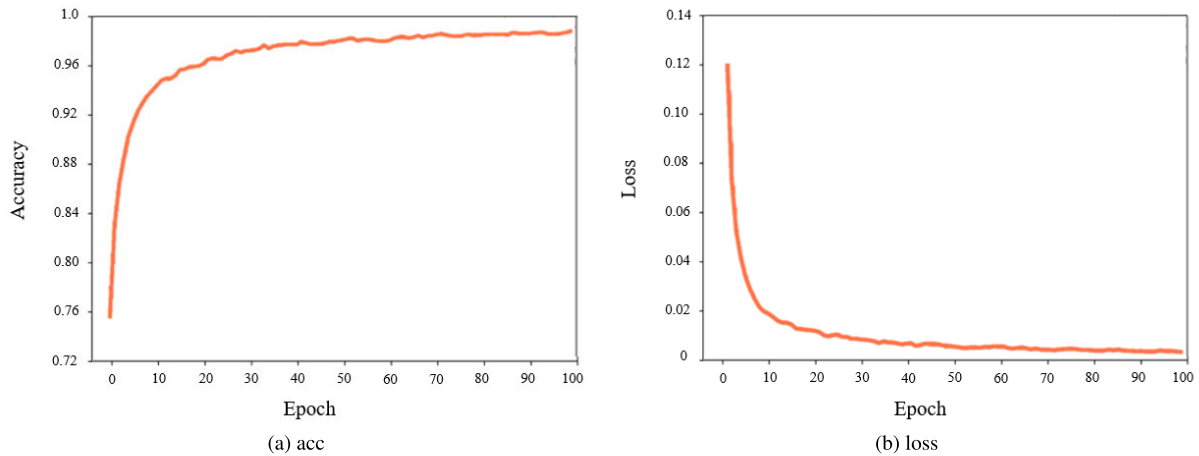
FIGURE 5. **Training curve of PSFNet on the training set.**

Secondly, the areas that contain classification weight in the image are different. Grad-CAM shows that ShuffleNet v2+PA has a larger area that contains classification weight in the image, which better covers the surface of the text. Meanwhile, the coverage of ShuffleNet v2+SE, ShuffleNet v2+CA, and ShuffleNet v2+CBAM is relatively smaller than that of ShuffleNet v2+PA.

Thirdly, different attention modules have different recognition capabilities for weakly textured images. We can see that the text content on fine-line patterns is easily disturbed by the background. The recognition of models embedding CA or CBAM modules does not seem to work well. While ShuffleNet v2+SE and ShuffleNet v2+PA have better performance in recognizing weakly textured images.

As a consequence, in combination with the model evaluation results of different ShuffleNet v2 in the dataset, it can be concluded that ShuffleNet v2+PA has better feature extraction capability than those of embedding other attention mechanisms.

D. ABLATION STUDY

We introduced the pixel attention module and focal loss function into ShuffleNet v2 respectively and conducted ablation experiments to verify the contribution of the improvement points proposed in this study to the network performance improvement. The comparison of the final training results is shown in Table 3.

It shows that ShuffleNet v2 itself has good performance and achieves a high recognition accuracy of 94.7% despite being a lightweight network. Attention mechanism

PA enhances feature characterization by generating attention coefficients for all pixels of the feature map, with 0.2% improvement in accuracy. However, when the focal loss was introduced, the accuracy of the model dropped to 93.8%. Finally, the pixel attention module and focal loss function help the improved model achieve an accuracy of 95.5% in the passport security feature dataset, improving by 0.8% compared to the baseline network with only a small increase in cost.

The test results showed that: 1) with an increase in the number of acceptable model parameters, the performance of the model combining the two improved strategies outperforms that of the model with one of the strategies alone; 2) Two improved strategies can be effectively integrated, and the improved model based on ShuffleNet v2 increases the accuracy of security feature recognition in complex context by 0.8 percentage points.

E. TRAINING RESULTS OF MODEL

Figure 5 shows the trend of loss and accuracy in 100 epochs of the training process. The loss starts to converge after 30 epochs and starts to stabilize after 50 epochs with a loss close to 0. The training accuracy starts to converge after 30 epochs and shows a stable state close to 99% after 60 epochs, the accuracy fluctuates from 98% to 99%.

In order to better evaluate the performance of the deep learning-based image classification model for passport security features, this experiment introduces the confusion matrix to reveal the classification performance of the improved
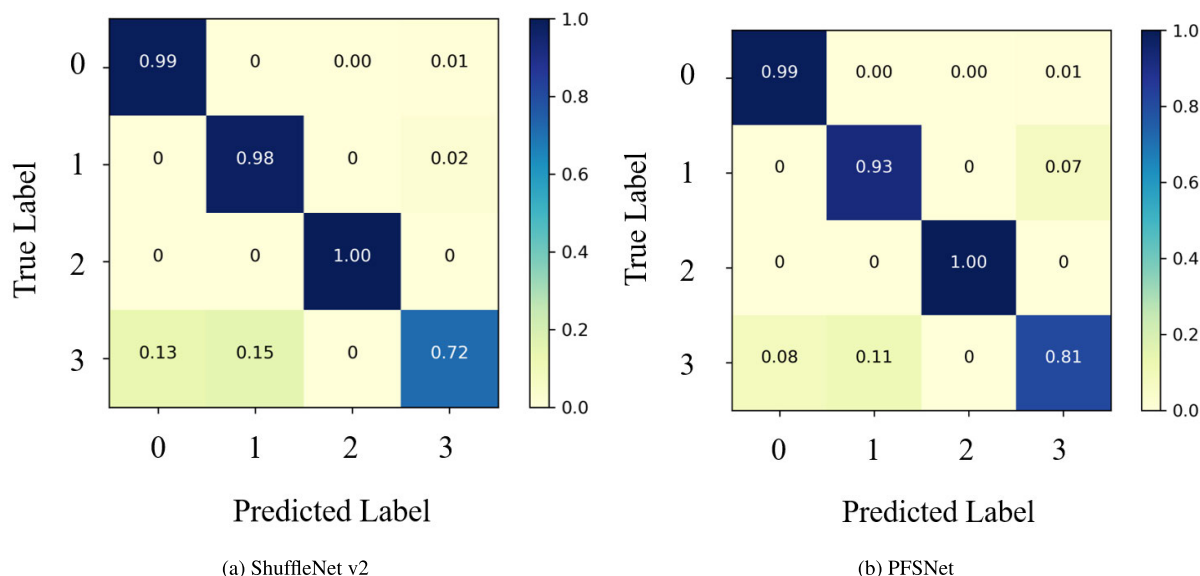
(a) ShuffleNet v2

(b) PFSNet

FIGURE 6. Confusion matrices of ShuffleNet v2 and PFSNet in the test set.

ShuffleNet v2 model. Figure 6 shows the recognition results of the passport security feature images test set before and after the improvement.

This study sets labels corresponding to 0, 1, 2, and 3 for Relief fine-line patterns, Intaglio printing, Laser engraving, and Relief embossing. The horizontal coordinate indicates the labels of the predicted passport security feature types, and the vertical coordinate indicates the labels of the real passport security feature types.

We can see that the number of images identified incorrectly by the original ShuffleNet v2 for label 3 is relatively high. The accuracy was only 72%, with 13% and 15% of Laser engraving images misclassified as label 0 (Intaglio printing) and label 1 (Relief embossing), respectively. However, PFSNet greatly improves the confusion between labels, This method increases the accuracy rate of label 3 from 72% to 81%, improving the accuracy of identifying the passport security feature test set effectively.

As PA can adaptively readjust the pixel-level features in all feature maps, PSFNet embedded with PA obtains better training results.

Relief embossing (label 1) and Laser engraving (label 3) were misclassified more often than other security features. Another 8% of Laser engraving (label 3) was misclassified as Intaglio printing label 1. To explain the cause of this phenomenon, we extracted a portion of the misclassified images for observation and analysis. Figure 7 shows two misclassified Laser engraving images. We can see that the text using Laser engraving looks raised, and most of the fonts are black in color. The three-dimensional effect of the font under the light is very outstanding. Under oblique light, relief embossing and intaglio printing also give the text a very visible raised effect, which is very similar to laser engraving text.
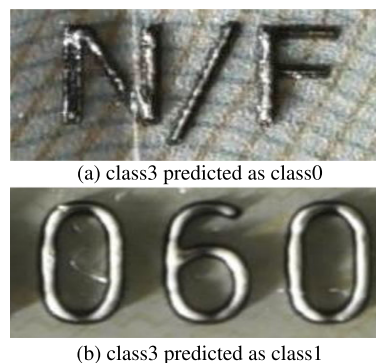

(a) class3 predicted as class0

(b) class3 predicted as class1

FIGURE 7. Incorrectly classified samples in the passport security feature identification experiment.

Therefore it is visually difficult to classify them accurately in a particular light and background.

To perform visualization tests of model recognition results, a set of passport security feature images were selected and these images were passed to the deployed PSFNet. Each image shows the results of the classification. As shown in Figure 8, the model is still able to accurately identify the category of the images in the case of complex background and blurred images, indicating that the model has a strong generalization ability. Therefore, PSFNet is an effective model for passport security feature classification which has a balance of speed and accuracy of recognition.

F. COMPARE WITH STATE-OF-THE-ART MODELS

There are a variety of lightweight models of convolutional neural networks for image classification, to illustrate the performance of the model proposed in this paper, we used several

TABLE 4. Comparison of training results of each model on the self-built dataset.

Model	Train ACC (%)	Train loss	Val ACC (%)	Val loss	Number of parameters	FLOPs	inference speed
ShuffleNet v2	98.4%	0.044	94.7%	0.21	1.26M	151.74M	0.00409
EfficientNet	98.1%	0.057	93.1%	0.25	4.01M	413.87M	0.00385
SqueezeNet	97.3%	0.076	94.0%	0.28	0.74M	732.74M	0.00401
MobilenetV2	98.1%	0.051	94.2%	0.22	2.23M	326.28M	0.00404
PSFNet	98.4%	0.083	95.5%	0.15	2.31M	203.12M	0.00369

FIGURE 8. Visualization of model recognition results achieved satisfactory results.

mainstream models for experimental comparison. We chose four classical deep learning models for comparative studies to compare our model and other machine models: ShuffleNet v2, EfficientNet, SqueezeNet, and MobilenetV2. These four models are common lightweight models in today's research.

The comparison findings are shown in Table 4. The accuracies of all five models in the training set are relatively similar. In the test set, the accuracy of PSFNet can reach 95.5%, which is more accurate than other models, and the loss of 0.15 is the lowest among all models. In addition, in terms of model size, the number of parameters of PSFNet is 2.31M, which is much smaller than the common CNN models. Among the five models listed in the table, the FLOPs of PSFNet are only 203.12M, which is only a little larger than that of the original ShuffleNet v2. While the same parameters of EfficientNet, SqueezeNet, and Mobilenet v2 are much larger than PSFNet.

In terms of inference speed, PFSNet's single-image test time is 0.00369s, which is the lowest among the five models. Therefore, we can see that PFSNet has an excellent performance in terms of testing speed and accuracy.

It can be seen that the model after introducing pixel attention and focal loss does not increase the parameters too much, while the accuracy is improved. So it proves that focal loss and pixel attention modules can help achieve better classification results in passport security feature image classification tasks.

V. CONCLUSION

Automated passport detection is an urgent problem that needs to be solved for future border management. Current classification methods rely too much on human involvement, and efficient and reliable automatic classification technology is an inevitable trend for social development. Applying artificial intelligence technology to passport security feature classification can improve classification efficiency and further reduce labor costs. In this study, we improve a lightweight neural network and the experimental results show that the embedded pixel attention module can improve the model accuracy with less cost. The recognition effect of introducing focal loss is better than the original model. The low consumption and high accuracy of passport security feature image classification are achieved. The algorithm in this study mainly considers single-label image classification, which is inadequate in recognizing multiple target objects of a single image. Our future work will focus on extending the passport security feature classification task to multi-label classification to achieve multi-object classification and recognition so that it can be applied to more passport security feature classification scenarios.

DECLARATION OF COMPETING INTEREST

The authors declare there are no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] S. Baechler, "Document fraud: Will your identity be secure in the twenty-first century?" *Eur. J. Criminal Policy Res.*, vol. 26, no. 3, pp. 379–398, Sep. 2020.

[2] E.-K. Prokkola and J. Ridanpää, "Border guarding and the politics of the body: An examination of the Finnish border guard service," *Gender, Place Culture*, vol. 22, no. 10, pp. 1374–1390, Nov. 2015.

[3] M. Husam, A. Kamel, D. F. Chalob, and Z. M. Abood, "Safe fingerprint analysis technique employed in passport verification for sustainable development," *J. Green Eng.*, vol. 10, pp. 6103–6114, Sep. 2020.

[4] M. Gariup and J. Piskorski, "The challenge of detecting false documents at the border: Exploring the performance of humans, machines and their interaction," *Int. J. Crit. Infrastruct. Protection*, vol. 24, pp. 100–110, Mar. 2019.

[5] G. Chiron, F. Arrestier, and A. M. Awal, "Fast end-to-end deep learning identity document detection, classification and cropping," in *Proc. Int. Conf. Document Anal. Recognit.* Cham, Switzerland: Springer, 2021, pp. 333–347.

[6] A. Maurya, S. Dubey, O. Manjrekar, A. Furtado, K. Arya, and P. Amith, "A system for verifying non-standard personal identity documents using deep learning," *Authorea Preprints*, Apr. 2022.

[7] T. D. Pham, Y. W. Lee, C. Park, and K. R. Park, "Deep learning-based detection of fake multinational banknotes in a cross-dataset environment utilizing smartphone cameras for assisting visually impaired individuals," *Mathematics*, vol. 10, no. 9, p. 1616, May 2022.

[8] K. Teymournezhad, H. Azgomi, and A. Asghari, "Detection of counterfeit banknotes by security components based on image processing and GoogleNet deep learning network," *Signal, Image Video Process.*, vol. 16, pp. 1505–1513, Jan. 2022.

[9] L. Zhao, C. Chen, and J. Huang, "Deep learning-based forgery attack on document images," *IEEE Trans. Image Process.*, vol. 30, pp. 7964–7979, 2021.

[10] D. Tong, "Watermark identification from currency applying deep learning techniques," Auckland Univ. Technol., Auckland, New Zealand, Tech. Rep., 2022.

[11] A. B. Centeno, "Analysis of background textures in banknotes and identity documents for counterfeit detection," Ph.D. dissertation, Departament de Ciències de la Computació, Universitat Autonoma de Barcelona, Barcelona, Spain, 2019.

[12] T. Kyrychok and O. Korotenko, "3D light interferometry investigation of ink layer formation during intaglio printing," in *Proc. 14th Int. Conf. Correlation Opt.*, Feb. 2020, pp. 246–254.

[13] H. Wang and L. Sun, "Materials and society: Functional chemicals and materials in banknote and security paper," *Accounts Mater. Res.*, vol. 2, no. 1, pp. 1–6, Jan. 2021.

[14] S. Kasman, "Impact of parameters on the process response: A Taguchi orthogonal analysis for laser engraving," *Measurement*, vol. 46, no. 8, pp. 2577–2584, Oct. 2013.

[15] M. Barta, "Travel documents and id cards counterfeiting in Slovak republic," *Zurnal Za bezbjednost I kriminalistiku*, vol. 2, no. 1, pp. 25–34, 2020.

[16] P. Poldrugac and E. Sop, "Technology for tactility in security printing," *TISKARSTVO* p. 67, Feb. 2011.

[17] R. Khallaf and M. Khallaf, "Classification and analysis of deep learning applications in construction: A systematic literature review," *Autom. Construction*, vol. 129, Sep. 2021, Art. no. 103760.

[18] B. Huang, H. Dou, Y. Luo, J. Li, J. Wang, and T. Zhou, "Adaptive spatiotemporal transformer graph network for traffic flow forecasting by IoT loop detectors," *IEEE Internet Things J.*, early access, Sep. 26, 2022, doi: 10.1109/JIOT.2022.3209523.

[19] S. Yang, H. Li, Y. Luo, J. Li, Y. Song, and T. Zhou, "Spatiotemporal adaptive fusion graph network for short-term traffic flow forecasting," *Mathematics*, vol. 10, no. 9, p. 1594, May 2022.

[20] F. Giannakas, C. Troussas, I. Voyiatzis, and C. Sgouropoulou, "A deep learning classification framework for early prediction of team-based academic performance," *Appl. Soft Comput.*, vol. 106, Jul. 2021, Art. no. 107355.

[21] B. Huang, G. Tan, H. Dou, Z. Cui, Y. Song, and T. Zhou, "Mutual gain adaptive network for segmenting brain stroke lesions," *Appl. Soft Comput.*, vol. 129, Nov. 2022, Art. no. 109568.

[22] T. Zhou, H. Dou, J. Tan, Y. Song, F. Wang, and J. Wang, "Small dataset solves big problem: An outlier-insensitive binary classifier for inhibitory potency prediction," *Knowl.-Based Syst.*, vol. 251, Sep. 2022, Art. no. 109242.

[23] E. Ledinauskas, J. Ruseckas, J. Marozas, K. Karlauskas, J. Terentjevas, A. Macijauskas, and A. Jursenas, "Automatic travel pattern extraction from visa page stamps using CNN models," 2021, *arXiv:2112.00348*.

[24] M. M. Abdelhameed, Y. A. Attia, M. S. Abdelrahman, and T. A. Khattab, "Photochromic and fluorescent ink using photoluminescent strontium aluminate pigment and screen printing towards anticounterfeiting documents," *Luminescence*, vol. 36, no. 4, pp. 865–874, Jun. 2021.

[25] S. G. MandeepKaurPurba and R. R. Gupta, "Differentiation of different blue inks in manipulated documents using VSC®-8000/HS: A comprehensive non-destructive technique," *Ann. Romanian Soc. Cell Biol.*, vol. 25, no. 6, pp. 5794–5801, 2021.

[26] M. K. Purba, S. Gupta, and R. R. Gupta, "VSC®-8000/HS—A robust tool for differentiation of overlapped black pen inks in fraudulent documents," Tech. Rep., 2021.

[27] M. Król, D. Kowalska, and P. Koscielniak, "Examination of Polish identity documents by laser-induced breakdown spectroscopy," *Anal. Lett.*, vol. 51, no. 10, pp. 1592–1604, Jul. 2018.

[28] M. J. Marques, R. Green, R. King, S. Clement, P. Hallett, and A. Podoleanu, "Non-destructive identification document inspection with swept-source optical coherence tomography imaging," in *Proc. Opt. Coherence Imag. Techn. Imag. Scattering Media IV*, Dec. 2021, pp. EW4A–6.

[29] H. Dou, J. Tan, H. Wei, F. Wang, J. Yang, X.-G. Ma, J. Wang, and T. Zhou, "Transfer inhibitory potency prediction to binary classification: A model only needs a small training set," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, Art. no. 106633.

[30] Y. Yuan, T. Quan, Y. Song, J. Guan, T. Zhou, and R. Wu, "Noise-immune extreme ensemble learning for early diagnosis of neuropsychiatric systemic lupus erythematosus," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 7, pp. 3495–3506, Jul. 2022.

[31] H. Huang, S. Zheng, Z. Yang, Y. Wu, Y. Li, J. Qiu, Y. Cheng, P. Lin, Y. Lin, J. Guan, D. J. Mikulis, T. Zhou, and R. Wu, "Voxel-based morphometry and a deep learning model for the diagnosis of early Alzheimer's disease based on cerebral gray matter changes," *Cerebral Cortex*, Mar. 2022, Art. no. bhac099, doi: 10.1093/cercor/bhac099.

[32] J. Qiu, G. Tan, Y. Lin, J. Guan, Z. Dai, F. Wang, C. Zhuang, A. H. Wilman, H. Huang, Z. Cao, Y. Tang, Y. Jia, Y. Li, T. Zhou, and R. Wu, "Automated detection of intracranial artery stenosis and occlusion in magnetic resonance angiography: A preliminary study based on deep learning," *Magn. Reson. Imag.*, vol. 94, pp. 105–111, Dec. 2022.

[33] Y. Song, T. Zhou, J. Y.-C. Teoh, J. Zhang, and J. Qin, "Unsupervised learning for CT image segmentation via adversarial redrawing," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 309–320.

[34] S. Kim, L. T. Nguyen, K. Shim, J. Kim, and B. Shim, "Pseudo-label-free weakly supervised semantic segmentation using image masking," *IEEE Access*, vol. 10, pp. 19401–19411, 2022.

[35] Z. Gao, H. Zhang, S. Dong, S. Sun, X. Wang, G. Yang, W. Wu, S. Li, and V. H. C. de Albuquerque, "Salient object detection in the distributed cloud-edge intelligent network," *IEEE Netw.*, vol. 34, no. 2, pp. 216–224, Mar. 2020.

[36] Z. Gao, C. Xu, H. Zhang, S. Li, and V. H. C. de Albuquerque, "Trustful internet of surveillance things based on deeply represented visual co-saliency detection," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4092–4100, May 2020.

[37] A. A. Jeny, M. S. Junayed, and S. T. Atik, "PassNet–country identification by classifying passport cover using deep convolutional neural networks," in *Proc. 21st Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2018, pp. 1–6.

[38] W. Zaaboub, L. Tlig, M. Sayadi, and B. Solaiman, "Neural network-based system for automatic passport stamp classification," *Inf. Technol. Control*, vol. 49, no. 4, pp. 583–607, Dec. 2020.

[39] L. Wang, M. Hu, D. Li, Z. Che, and X. Zhang, "Security thread detection in passport using improved template matching," in *International Forum on Digital TV and Wireless Multimedia Communications*. Berlin, Germany: Springer, 2017, pp. 78–90.

[40] J. Kim, "Recognition of passport MRZ information using combined neural networks," *J. Korea Soc. Digit. Ind. Inf. Manag.*, vol. 15, no. 4, pp. 149–157, 2019.

[41] Y. Liu, H. James, O. Gupta, and D. Raviv, "MRZ code extraction from visa and passport documents using convolutional neural networks," *Int. J. Document Anal. Recognit.*, vol. 25, no. 1, pp. 29–39, 2022.

[42] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 116–131.

[43] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 13713–13722.

[44] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image super-resolution using pixel attention," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 56–72.

[45] H. Peng, X. Chen, and J. Zhao, "Residual pixel attention network for spectral reconstruction from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 486–487.

[46] S. Gao, J. Zhu, and Y. Yang, "CP-Net: Channel attention and pixel attention network for single image dehazing," in *Proc. Int. Conf. Pioneering Comput. Scientists, Eng. Educators*. Cham, Switzerland: Springer, 2020, pp. 577–590.

[47] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[48] T. F. Romdhane, H. Alhichri, R. Ouni, and M. Atri, "Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103866.

[49] G. S. Tran, T. P. Nghiem, V. T. Nguyen, C. M. Luong, and J.-C. Burie, "Improving accuracy of lung nodule classification using deep learning with focal loss," *J. Healthcare Eng.*, vol. 2019, pp. 1–9, Feb. 2019.

[50] K. Pasupa, S. Vatathanavaro, and S. Tungjitnob, "Convolutional neural networks based focal loss for class imbalance problem: A case study of canine red blood cells morphology classification," *J. Ambient Intell. Humanized Comput.*, early access, pp. 1–17, Feb. 2020.

[51] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?" 2016, *arXiv:1611. 07450*.

[52] Y. Liu, Z. Zhang, X. Liu, W. Lei, and X. Xia, "Deep learning based mineral image classification combined with visual attention mechanism," *IEEE Access*, vol. 9, pp. 98091–98109, 2021.

ZHIZHE LIN is currently the Dean of the Office of Emergency Management, Shantou Central Hospital. Her research interests include computer vision, machine learning, and pattern recognition.

JIAFENG XU is currently an Assistant Professor with the Immigration Management College, China People's Police University (Guangzhou). His research interests include computer vision, machine learning, and document detection.

DAWEI JIA is currently an Assistant Professor with the Immigration Management College, China People's Police University (Guangzhou). His research interests include computer vision, machine learning, and digital security technology.

TENG ZHOU is currently an Assistant Professor with the Department of Computer Science, Shantou University, China, and also works as a Research Fellow at the Key Laboratory of Intelligent Manufacturing Technology, Shantou University, Ministry of Education. His research interests include intelligent transportation systems and machine learning.

• • •