Computer Science and Information Systems 22(1):219–238 https://doi.org/10.2298/CSIS240227001X

Medical Images Anomaly Detection for Imbalanced Datasets with Multi-scale Normalizing Flow

Yufeng Xiao[1,2,3], Xueting Huang[1,2,3],Wei Liang[1,2,3,⋆], Jingnian Liu[1,2,3], Yuxiang Chen[1,2,3], Rui Xie[1,2,3], Kuanching Li[1,2,3,⋆] and Nam Ling[4]

[1] School of Computer Engineering and Science, Hunan University of Science and Technology, Xiangtan 411100, China
[2] Sanya Research Institute, Hunan University of Science and Technology, Sanya 572024, China
[3] Hunan Key Laboratory for Service Computing and Novel Software Technology, Xiangtan 411201, China
hnxiaoyf@hnust.edu.cn
HXuetingmail@163.com
wliang@hnust.edu.cn
jingnianl@mail.hnust.edu.cn
chenyuxiang@hnust.edu.cn
xierui54@mail.hnust.edu.cn
aliric@hnust.edu.cn
[4] Department of Computer Science and Engineering, Santa Clara University, USA
nling@scu.edu

Abstract. Due to the substantial feature extraction and end-to-end learning capability, deep learning has been widely used in intelligent medical image detection. However, amount of parameters in these models relies on the number of labeled training data, which influences the performance. Due to this reason, we propose a novel unsupervised medical image detection model named Multi-Scale Normalizing Flow (MS-NF). First, a fusion backbone network is applied to extract the multi-scale feature maps, which capture the different scale features of the anomalies. Second, normalizing flow transfers the abnormal distribution into the normal distribution hidden in the latent space, which is used for anomaly detection. To further improve the detection performance, channel and spatial convolutional attention mechanisms are integrated to make the model focus on the anomalous region by a shared network. Experimental results obtained on brain tumor MRI and ISIC2018 datasets show that MS-NF improves the pixel-level AUC index by 9% compared to existing medical image detection models, also performing well on small-scale data with efficient training and inference.

Keywords: Anomaly Detection, Unsupervised learning, Attention Mechanisms, Normalizing Flow, Medical Images.

1. Introduction

Medical image anomaly detection through image processing and deep learning can automatically detect anomalies and localize abnormal regions. It has been widely applied in assisted diagnosis, treatment, condition assessment, and medical decision support. In the

⋆ Corresponding authors

past decade, diverse machine learning models has been introduced to establish anomaly detection models, such as SVM [24], Decision Trees, and KNN. These models take the handcrafted features as the input, which heavily relies on expert knowledge. Moreover, the shallow structure of machine learning limits their capability to learn the deep representations hidden in the medical images.

Unlike traditional machine learning, Deep Learning (DL), which consists of a large number of hidden layers and nodes, can distill the complex features hidden in the data. Since it has successfully been applied in the Internet of Things [20, 3, 14], Blockchain[19, 39, 41], Quality of Service [17, 18], data recovery [16, 10], DL is also introduced to fuse low- and high-dimensional image features by jump-joining in the field of the medical image detection. Taking advantage of learning ability, it can effectively extract low- and high-resolution features of medical images when containing noise, blurred boundaries, and so on**??**. Indeed, automatic feature extraction can improve the effectiveness and performance of the training model.

Undoubtly, medical image detection is a supervised task, while DL is a data-driven model. Their performance heavily depends on the number of the training dataset. However, it is time-consuming to obtain the amount of high-quality labeled medical images that rely on professional knowledge. In other words, the procedure of the annotation is subjective and inconsistent. Worse, abnormal medical images are usually scarce compared to the normal samples. Therefore, training the medical images often faces the challenge of data imbalance in real-world application scenarios. It means that deep learning mainly learns the hidden representations of the normal samples, making it difficult to recognize the abnormal samples. Meanwhile, the location of the lesion is uncertain that has a variety of shapes and sizes.

To avoid the dependence on the labeled data, a diversity of unsupervised models are introduced to process the medical image detection, such as variational Autoencoder (VAE) [12], generative adversarial network (GAN) [4] and their variational models [11, 13]. The core idea of these models is to transform the abnormal region into a normal medical image which can generate a new abnormal medical image. The generated image is optimized by the differences between it and the original image. Therefore, the augmented medical images are helpful for lesion detection and localization. However, the performance is still affected by the inherent shortage of the generative models. For VAE, the high-dimensional data is mapped into the low-dimensional hidden feature space that may cause information loss. When performing reconstruction processing, the generated image is usually blurred, which harms medical image detection. Different from VAE, GAN directly models the distribution of the data. Nevertheless, it faces the challenges of training instability, gradient disappearance, and mode collapse. Denoising Diffusion Probabilistic Models (DDPM) is another popular generative model that generates high-quality samples without adversarial training. The self-encoder component in the DDPM denoises and restores the anomalous medical images corrupted by Gaussian noise. Since the denoising processing is carried out at each step, the model converges slowly.

Another common medical image detection method is to learn the distribution of normal data using deep learning. The anomalies are identified by distinguishing whether the distribution satisfies the normal data. It comprises two components: the feature extraction network and the distribution estimation network. The feature distribution of normal data is learned using deep convolutional networks, such as Resnet or visual architecture Vit based on transformer [36]. Then, normalizing flow is implemented to estimate the feature distribution likelihood of the sample data, which is used for classification. Though this method has widely been applied in industrial anomaly detection, though seldom used in the medical field.

Taking advantage of the coupling layers, a novel inverse neural network is established to estimate the distribution of the abnormality. Specifically, a backbone is used to extract the feature maps in terms of different scales, which are split into two parts. A trained normalizing flow takes them as input and splices them into the output matrix. According to the matrix, the likelihood probability between anomalous and healthy data is calculated to recognize the anomalies. Several public datasets, such as ISIC 2018 and Brain MRI, are chosen to evaluate the effectiveness of the proposed model. The contributions of this work are summarized as follows.

- A novel unsupervised model with normalizing flow is proposed for medical image detection, which can automatically identify the differences between normal and abnormal data. The differences are used to generate new abnormal medical images. These generated images are implemented to train the model, which can relieve the dependence on abnormal medical images.
- The proposed model integrates the convolutional attention mechanism and normalizing flow, which makes it easy to focus on the critical features. It can improve the detection performance.
- Experiments conducted on several public datasets show that the proposed model is superior to other state-of-the-art models.

The remainder of this article is organized as follows. The related work is introduced in Section 2. The methodology of the proposed model is described in Section 3. The experiments are discussed in Section 4. Concluding remarks and future directions are drawn in Section 5.

2. Related Work

Deep learning has been widely used to learn the features hidden in medical images. To relieve the dependence on the labeled data, unsupervised models are introduced to apply in medical image detection. They can roughly divided into two categories, including feature- and reconstruction-based models.

2.1. Feature-based Unsupervised Models

Feature-based models usually aim to perform feature extraction and distribution transformation. They can learn the critical information hidden in the latent feature space used to detect anomalies. For example, some pre-training models, such as AlexNet, VGG, which train deep neural networks on the large-scale dataset Imagenet are widely implemented to extract features from the medical images. Compared with AlexNet, VGG, which has a deeper network structure and smaller convolutional kernels, achieves better performance[15]. However, it faces the challenges of gradient vanishing, gradient explosion, and degradation. He *et al.* proposed a novel model Resnet by introducing the structure of residual connection to convolutional neural networks. It can effectively solve these problems by selecting the operation of skipping part of the layers [9]. Most of the work, anomaly detection based on unsupervised models, takes Resnet as the backbone network to extract features[29]. Some researches [21] use transformers-based visual networks Vit, DeiT, and their variations for unsupervised anomaly detection. Vit [6] and DeiT [34] use a self-attentive mechanism to extract features with better migration capabilities. Transferring learning across different visual tasks gives it better generalization to deal with multi-tasks. In addition, Vit and DeiT are more adaptable to large-scale datasets and robust to changes in data augmentation than other state-of-the-art models.

For distribution transformation, the mean and variance of normal features are used to model the normal distribution. When the probability distribution of the input data significantly deviates from the true distribution, it is recognized as an anomalous medical image. Normalizing Flows (NF) is a generative model, consisting of a series of invertible transformations[28]. It can learn probability distributions by mapping one distribution to another one. Combining these mappings forms a complete flow that can transform a complex distribution into a simple probability distribution. To address the problem of the poor efficiency of INN training and inference, Dinh et al. proposed Real-NVP by using reversible, local affine transformations to transform regularized streams of data distributions, which can improve computational efficiency[5]. Rudolph et al. presented Differnet, which accomplishes image detection through likelihoods provided by Normalizing Flow on multi-scale image features with multi-transform evaluation [30]. Compared to other models, Differnet requires fewer training samples. These works show promising results in the field of anomaly detection. Existing methods are based on reconstruction, and few works have applied them to medical image detection.

2.2. Reconstruction-based Methods for Generative Modeling

Variational Autoencoder(VAE) contains an encoder and a decoder. The encoder aims to learn the features hidden in the latent space and the decoder reconstructs the input data according to the latent features. Van et al. proposed a discretized VAE with a code book approach to learning the data distribution instead of VAE. However, it cannot generate high pixel-level samples[35]. Consequently, Razavi et al. introduced a hierarchical structure to propose VQ-VAE2[27]. It extends the number of discrete codebooks from the original thousands to millions that can capture the local features of the data in detail. Esser et al. incorporates a transformer structure, which generates millions of pixels[7]. In addition, the autoencoder is used to model the distribution of the normal data for medical image detection. The sample deviates from the distribution of the normal data that can be classified as abnormal data. Zimmerer et al. proposed a combination of density-based and reconstruction-based anomaly detection models that do not require labeled data and allow for anomaly scoring and localization of samples[43]. However, using VAE to reconstruct, the re-generated image is blurred, and anomalous regions may be retained. Nevertheless, VAE has been gradually replaced by subsequent competitive models.

The training of GAN is based on a set of neural networks that fight against each other. The generator generates the new data, which is discriminated the true or false by the discriminator [4]. The generator then adjusts its parameter weights according to the classification results of the discriminator to generate more realistic images. Based on [4], Radford et al. proposed Deep Convolutional GAN (DCGAN), which uses Convolutional

Neural Networks (CNNs) as generators and discriminators to produce higher-quality images. The experimental results show better image synthesis results[26]. Arjovsky et al. proposed Wasserstein GAN (WGAN), which uses Wasserstein distance to measure the difference between generated samples and real samples. It can overcome the instability when training GANs, which perform better image generation and speech synthesis [8]. Zhu et al. introduced CycleGAN to image translation tasks, which converts an image from one domain to an image from another domain [42]. Compared to the previously mentioned methods, CycleGAN does not require training data pairs and can be trained without matching data. Karras et al. proposed a Style-Based Generative Adversarial Network (StyleGAN), which uses a new generator architecture, including a style network and a generative network, to generate higher-resolution and more realistic images. Brock et al. proposed BigGAN, using more extensive networks and more sophisticated training strategies to generate higher-quality images, and has achieved leading results in various image-generation tasks[2].

GAN and its derived models are influential in anomaly detection. The generator can simulate the distribution of the real data. Therefore, if the input data differs significantly from the data generated by the generator, it can be considered anomalous data. Schleg et al. proposed AnoGAN, which is the first GAN-based anomaly detection method [32]. The normal sample distribution is obtained by trained DCGAN to find the abnormal region of the sample. To address the slow model speed of the technique in [32], Schleg et al. proposed an unsupervised anomaly detection method using GANS called F-AnoGAN [31], to replace the iterative optimization used in AnoGAN with a fast approximation algorithm based on Wasserstein distance. This approximation generates a latent representation of the normal data and a reconstruction loss for normal and abnormal samples. The reconstruction loss is used to score each sample and determine whether it is normal or abnormal. Bhatt et al. uses progressive GANs to improve the resolution of detecting anomalies, allowing for the processing of more detailed images [1]. However, the above adversarial generative network-based approach still suffers from training instability and difficulty in capturing the complete data distribution of the image.

Wyatt et al. first applied diffusion modeling to the field of anomaly detection and achieved good detection results in industrial, medical image detection, and other areas [38]. A more extensive range of anomalous regions can be recovered by replacing Gaussian noise with simple noise. Teng et al. applied a score-based model to visual defect detection [33]. Pinaya et al. proposed an encoder model for anomaly segmentation of medical images [23].

MRI images usually are low quality and have poor contrast due to image projection and laminar imaging processes. The boundaries between biological tissues are often blurred and difficult to detect, which makes it difficult for generative reconstruction methods to accomplish the task of pixel-level image segmentation. In contrast to the generative model, the proposed MS-FLOW can estimate the exact likelihood value. Therefore, it can accurately detect Out-Of-Distribution (OOD) samples. Moreover, it does not require a large amount of data for training and is more stable in training, with shorter convergence time and faster inference.

3. Methodology

To relieve the dependence on the label information and achieve a good performance, we propose a novel unsupervised model, MS-NF, which can detect and locate anomalous regions in medical images. The framework of the MS-NF is shown in Fig. 1, which consists of two components, including a feature extraction network and a distribution transformation network. Concretely, the first three residual blocks of Resnet18 are considered the backbone for feature extraction, whose outputs are fed into the distribution transformation network. Then, a normalizing flow transforms the feature maps into the latent space, where convolutional operation preserves the 2-dimensional information of the medical image. Moreover, a convolutional attention mechanism is applied to make the model more focused on the critical region of the medical image. This section details the proposed model's principles and presents the training and inference procedure.

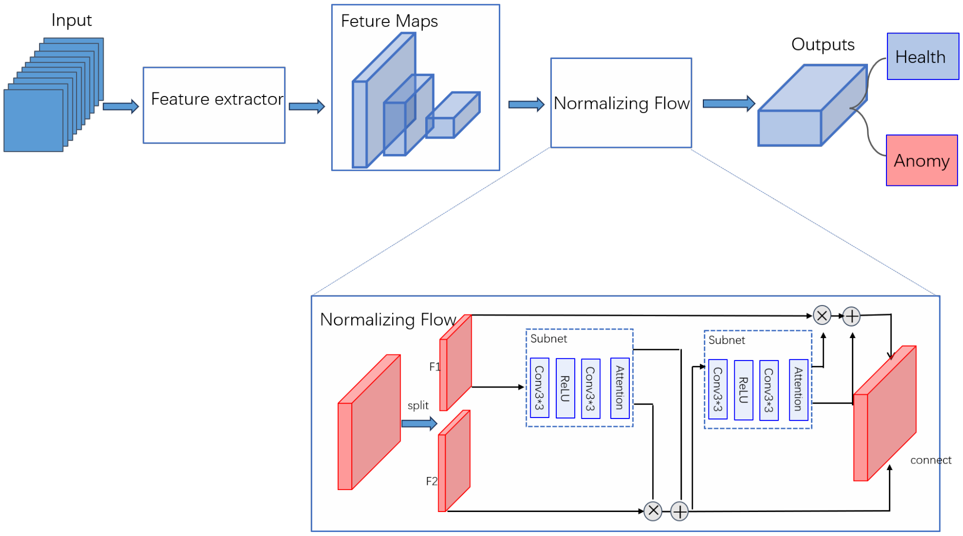

Fig. 1. The detailed view of MS-FLOW

3.1. Feature extraction network

Given a medical image, we denote it as $x \sim p(x)$, where x belongs to $X(x_1, x_2......)$. The backbone network is formed by deep convolutional neural network, such as Resnet or visual transformer-based network model Vit. Since these pre-training models is optimized by amount of images, it can be used to extract the critical features hidden in the medical images. Assuming ex: $X \to F$, it transforms the original information hidden in the image x to the high-level feature representation. There are some lesions or tumors in the abnormal regions, which have a diversity of characteristics in different organ. Therefore, it is important to extract the global information of the medical images. In this paper, we use the residual block of Resnet18 to extract features, which can learn the local and global information. And the last last layer of each residual block is implemented to capture the connection between local and global features. These features are fed into normalizing flow later.

3.2. Distribution transformation network

This article takes normalizing flow as the distribution transformation network, which transforms the image feature F into the latent space L. This is done to compute exactly the distributional likelihood for each feature element $f \in F$. Since the training consists entirely of normal samples, in the inference phase, the likelihood of the abnormal information in the medical images will be mapped to the edges of the distribution. At this point, a threshold is set to distinguish between abnormal and normal. The process of normalizing flow consists of a series of stacked invertible differentiable functions. The process of going from the characteristic distribution F to the Gaussian distribution L undergoes some series of flow changes. There is a one-to-one mapping relationship between the image features and the corresponding latent space features, and it can be expressed as $T : F \to L$. In other words, they are reversible. Assuming the processing stack is a series of transformation functions, it can be expressed as.

$$L = T_1(F) \circ T_2 \circ T_3 ... T_K. \tag{1}$$

Therefore, the anomalies that deviate from the distribution of the normal data can be restored to the original image according to the stacking invertible functions.

$$F = T_K(L) \circ T_{K-1} \circ T_{K-2} ... T_1, \tag{2}$$

Setting $l = T_\theta(f)$, the following expression for calculating the likelihood of the characteristic distribution can be obtained by applying variable substitution to Equation. 2:

$$\log p_\theta(f) = \log p_\theta(l) + \log \left| \det \left(\frac{\partial l}{\partial f} \right) \right| \tag{3}$$

Due to the properties of the flow-varying bijection, Equation.4 can be converted to Equation.5 by the chain rule:

$$log p_\theta(f) = \log p_\theta(l) + \sum_{i=1}^{k} \log \left| \det \left(\frac{\partial T_k(l)}{\partial T_{k-1}(l)} \right) \right| \tag{4}$$

The exact likelihood for the sample features is computed by transforming f into l. The parameters of the target distribution are optimized by maximizing the log-likelihood, where transformation network T is implemented by normalizing the flow component. The construction of the normalizing flow component is shown in Fig. 1. Therefore, the distribution transformation network stacks several normalizing flow components. Moreover, we introduce coupling layers according to [23, 40]. Concretely, the input data x is first separated into two parts x_1 and x_2. Transformations are respectively performed on x_1 and x_2 using normalizing flow. When the 2-dimensional flow layer transforms the 2-dimensional image into 1-dimensional space, it may occur the loss of medical image information may make the distribution transformation unstable. Consequently, 3×3 convolutional kernel is used to retain the 2-dimension information. The intermediate quantity y_1 is obtained by combining x_1 to x_2 using an invertible transformation.

$$y_1 = e^{sub_1(x_1)} \odot x_2 + sub_2(x_1) \tag{5}$$

Then, continue to perform transformations on y_1 using subnet. Finally, y_1 and y_2 are combined to get the final output tensor y:

$$y_2 = e^{sub_1^{'}(y_1)} \odot x_1 + sub_2^{'}(y_1) \tag{6}$$

$$y = concat(y_1, y_2) \tag{7}$$

where sub represents the normalizing flow component consisting of convolutional and attention layers. For the separation of x and y_1, it is performed on the channel. Respectively, the symbols $\odot$ and $+$ denote multiplication and addition along the feature direction.

3.3. Convolutional Attention

Medical images suffer from problems such as blurred boundaries and unclear imaging. In this work, to enable the NF module to focus more on the image region of interest for the detection task and thus improve the detection ability of the model. Therefore, the attention mechanism based on the convolutional block is embedded between the convolutional layers [37]. The detection performance is improved without significantly increasing the number and complexity of model parameters. Convolutional attention consists of two parts: channel attention and spatial attention. The intermediate feature map F, obtained by the NF module subnet, enters the channel attention layer and the spatial attention layer successively in order. The entire process can be summarized as follows:

$$F_i = A_c(F) \otimes F \tag{8}$$

$$F_o = A_s(F_i) \otimes F_i \tag{9}$$

The channel attention layer focuses on the information that is meaningful in the features. Through average pooling and maximum pooling, two feature vectors are generated and fed into a shared network. This shared network is a Multi-Layer Perceptron (MLP) with a hidden layer. Once the two feature vectors have passed through the forward network, the features are then merged for each element at the same location:

$$A_c\left(F\right) = \sigma\left(MLP\left(AvgPool\left(F\right)\right) + MLP\left(MaxPool\left(F\right)\right)\right) \tag{10}$$

Spatial attention focuses on where the features have the most information. Two feature vectors will be generated by average pooling and maximum pooling operations. The convolution operation is performed on these two feature vectors to get the spatial attention map:

$$[A_s\left(F\right) = \sigma\left(conv^{7\times 7}\left(\left[AvgPool\left(F\right); MaxPool\left(F\right)\right]\right)\right)] \tag{11}$$

where σ represents a sigmoid function, while $conv$ represents the convolution operation with a kernel size of 7×7. By combining channel attention and spatial attention, we capture inter-channel dependencies and intra-channel spatial relationships to enhance the performance of the proposed model. The two attention mechanisms can be used in parallel or a sequential order. In this work, we adopt the arrangement proposed in [37], which integrates channel attention and then incorporates spatial attention.

We use the following algorithm 1 to construct our MS-NF module as a sub-network of the feature transformation module: the details are as follows: first input the backbone features f, the convolution size $kernel\_size$, and the $reduction$ that controls the proportion of dimensionality reduction in the channel's attention mechanism. Next, the features f at the different scales are input to each nf module in order. For each module, a sequence of reversible neural networks is created. The feature f will be successively processed by the 2D convolution, the RELU, the 2D convolution, and the convolutional attention layer, and the transformed feature probabilities will continue to be sent to the next nf module until the end of the loop.

Algorithm 1: Normalizing flow with attention

```
Input:Image features $f$, kernel_size, reduction
Output:$y$
Procedure:
  Initialize()
    $c1 \leftarrow$ conv2d(in_channels, hidden, kernel_size = cc)
    $ac \leftarrow$ relu()
    $c2 \leftarrow$ conv2d(hidden, $out$, kernel_size = ...)
    $att \leftarrow$ Attention(reduction = 16)
  for $nf_i$ in $NF_n$ :
    $y \leftarrow$ conv2d($F$)
    $y \leftarrow$ relu($y$)
    $y \leftarrow$ conv2d($y$)
    $y \leftarrow$ att($y$)
  end for
```

3.4. Training and Inference

For normalizing flow, the model is trained entirely on normal medical images. We would like to get the exact likelihood value of the sample feature distribution, but it isn't easy to solve pf directly. So the log-likelihood of the feature $f \in F$ is estimated by, *i.e.*, the potential space, as shown in Equation 12. The method aims to find suitable parameters to maximize the probability density function of the feature f extracted by the backbone. Therefore, the loss is defined as follows:

$$loss = -\log p_{\theta}\left(f\right) = -\log p_{\theta}\left(l\right) - \log\left|\det \frac{\partial l}{\partial f}\right| \qquad (12)$$

Algorithm 2: Training and Inference

```
**Input:**Image image $x$
**Output:**$y$
Procedure:
  Feature_exactor ← create_model(resnet18,pretrained=False,features=true,
  out_indices=[1,2,3,],in_chans=1)
  $f$ ← Feature_exactor(x)
  output,log_jac_dets ← nfblock(f)
  $loss \leftarrow \frac{\|l\|_2^2}{2} - \log\left|\det \frac{\partial l}{\partial f}\right|$
  If not training
    anomap mean(output**2,dim=1)
  return anomap
```

In the inference stage, because the anomalous features never appear in the training data, the probability density function of the anomalous features obtained by the model is lower than that of the normal data. When the normalization flow transforms the abnormal features into the latent space, the data distribution lies in different intervals. When T_{NF} transforms the anomalous features into the latent space, a different distribution is obtained. The likelihood of the features can be used as an anomaly score in detection. We set a threshold τ, for a normal image. The likelihood on the feature map is more significant than τ:

$$Ano = \begin{cases} 1, p_L\left(T_\theta\left(F\left(x\right)\right)\right) > \tau \\ 0, p_L\left(T_\theta\left(F\left(x\right)\right)\right) < \tau \end{cases} \tag{13}$$

When $Ano = 1$, it means there is an anomaly in the current medical image.

The following is our training and inference algorithm. It is done as follows. First, the backbone network of resnet18 is constructed for feature extraction, and the loss function for training is obtained according to Eq. (10) to optimize the parameters. If the model is in the inference stage, the likelihood of the probability distribution of each pixel point is calculated, and then an anomaly map is obtained for anomaly detection.

4. Experimental Results

In this section, we evaluate the performance of the proposed model MS-FLOW, which is compared with other state-of-the-art methods collected in the literature [31] [38] [40]. In addition, ablation experiments are performed to verify the influence of each model component.

4.1. Datasets

Normalizing flow learns the distribution of the sample data, while the abnormal data are outside the distribution. The training process is performed entirely on normal datasets. The training set is divided into two parts, including brain MRI impact and dermatologic images. The brain MRI images is NFBS dataset [25]. The test set is a neuroimaging dataset of brain tumor patients (NIBT) [22]. NFBS dataset consists of 125 T1-weighted MRI (magnetic resonance images) with the size of $256 \times 256 \times 192$. In this paper, only single slices of the transverse axis are considered, and a single-channel image of size 256×256 is intercepted. Since most of the critical information of brain MRI images is concentrated in the middle of the brain, we randomly selected the slices from 40 to 80. Since the edge portion of the image has a black background, we perform center cropping on each medical MRI image to make the model focus on the brain region. Small rotations and translations are applied to the images to enhance image diversity. The input is uniformly resized and normalized to improve training stability. A neuroimaging dataset, brain tumor patients (NIBT) [22], is chosen as the test set, which consists of T1, T2, fMRI, and DTI MRI images for 22 brain tumor patients. The size of the images in this dataset is $256 \times 256 \times 192$. For evaluation, we also consider transverse slices. Since most brain tumors occurred in the middle, we discarded the beginning and the end. Therefore, the slices of the image are randomly selected in the range of 140 to 200 slices. Then, data enhancement operations, such as center cropping, resizing, and normalization, are performed on each medical image. For the Brats2020 dataset, we take slices from the center region. For slices with mask 0, it is treated as normal, while others are treated as abnormal.

The other part is a dataset (Task 3) from the ISIC2018 challenge, which is related to the classification and localization of dermatological images. Task 3 contains seven categories in total. We consider the nevus type as normal data and the other types as abnormal data. The training set contains 8224 nevus images and the test set contains 1514 images of other types. The data is normalized and resized to 256×256 size.

4.2. Implementation Details

We use Pytorch to implement the proposed model. The construction of the inn network is completed through the FrEIA library. Resnet18 is treated as backbone, which obtains multi-scale feature maps by fusing the outputs of the first three residual blocks. We do not choose the pre-training model due to the difference between MRI and imagenet images. The input of the model is unified as a single-channel brain MRI image with a size 256×256. The three scales' feature map channels are 64, 128, and 256, respectively. The distribution transformation model consists of 15 NF blocks. Each sub-network consists of a 3×3 2D convolutional network, a relu layer, and a convolutional attention layer (CBAM). The construction of CBAM is completed by the fightcv_attention library, and the reduction parameter that specifies the compression ratio of the number of channels is set to 16. In addition, the dimension of the hidden feature learned is set as 128.

During training, the images are subjected to random rotations ranging from -3 degrees to $+3$ degrees and random translations of width 0.02 and height 0.09. The batch size is set to 15. The Adam optimizer is used as the parameter optimizer with a learning rate of 10^{-3}. The weight decay value is set to 10^{-5}. During inference, the performance of the proposed method is evaluated against other comparative methods using the Area Under the Receiver Operating Characteristic Curve (AUROC). To evaluate the performance of the model on the test data set, we used metrics such as the area under the receiver operating characteristic curve (AUROC), accuracy (Acc), recall, average precision (AP), and F1.

4.3. Quantitative Results

We quantitatively compare the proposed model with the remaining three deep learning-based anomaly detection methods. They contain a representation-based method, Fast-flow [40], and two image reconstruction-based methods, f-AnoGan [31] and AnoDDPM [38]. For testing, we evaluate the proposed model and the compared models on two public datasets, including Brain MRI and ISIC2018. Table 1 collects the results of the quantitative analysis of the proposed model and the remaining three methods on the brain MRI dataset. Table 2 demonstrates the results of the quantitative analysis of our method and the remaining three methods on the ISIC2018 dataset.

Table 1. Quantitative analysis for different methods on Brain MRI

	Auc	Acc	Recall	Ap	F1
Ours	0.91	0.84	0.78	0.45	0.42
f-AnoGan[31]	0.84	0.75	0.81	0.10	0.14
AnoDDPM[38]	0.74	0.98	0.61	0.67	0.64
Fast-flow[40]	0.72	0.73	0.72	0.24	0.32

Table 2. Quantitative analysis for different methods on ISIC2018

	Auc	Acc	Recall	Ap	F1
Ours	0.81	0.72	0.72	0.70	0.67
f-AnoGan[38]	0.78	0.67	0.69	0.68	0.66
Fast-flow[40]	0.72	0.68	0.67	0.62	0.64
Autoencoder	0.70	0.62	0.60	0.65	0.65

Table 3. The performance of MS-NF with different backbones selected on the Brain MRI

Conv	Backbone	Metric Values				
		Auc	Acc	Recall	Ap	F1
3×3	Resnet18	0.91	0.84	0.78	0.41	0.39
	WideResnet50	0.73	0.68	0.70	0.09	0.18
	Cait	0.66	0.51	0.81	0.07	0.14
	Deit	0.65	0.35	0.90	0.08	0.12
3×3 and 1×1	Resnet18	0.77	0.73	0.81	0.10	0.23
	WideResnet50	0.54	0.46	0.68	0.05	0.11
	Cait	0.66	0.51	0.81	0.07	0.14
	Deit	0.65	0.35	0.90	0.08	0.12

We select the threshold with the slightest difference between True Positive Rate (TPR) and False Positive Rate (FPR) on the ROC curve as the accurate value for calculating other evaluation criteria. Figure 2 shows the ROC curve of this method and the other three methods on the NIBT dataset. Our method is superior to other compared models.

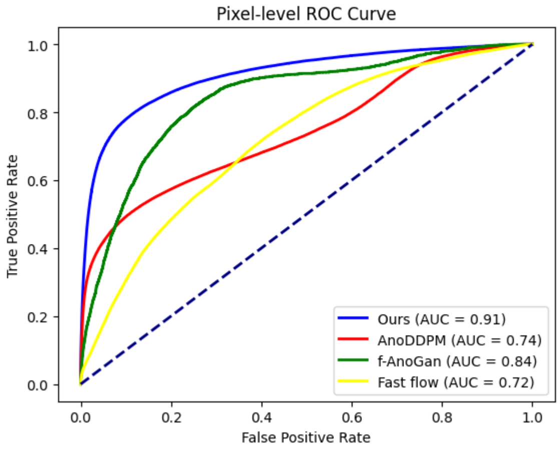

Fig. 2. Demonstration of ROC curves for different methods

MS-FLOW achieves the best results in the AUROC metric. It also performs well in ACC, recall, and so on. By transforming the feature distribution to a standard normal distribution through the NF module, abnormal data deviates from the normal distribution, completing the localization. Figure 3 shows the localization effect of MS-FLOW on brain MRI data. Columns 1 to 5 are abnormal data. The latter column is a normal brain MRI scan. For the tumor portion, MS-FLOW localizes its abnormal portion.

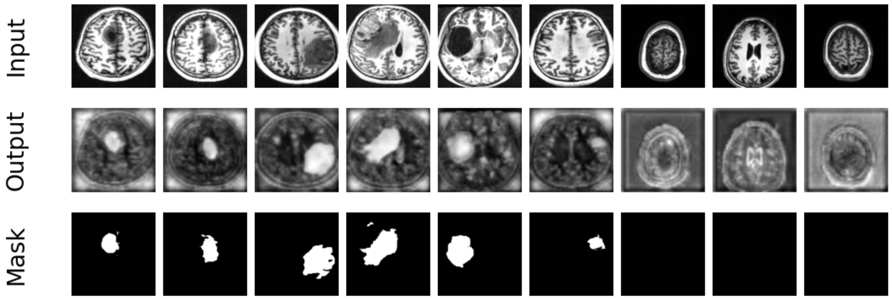

Fig. 3. Detection effect diagram. Output represents the model-predicted images, and Mask represents the ground truth

f-Anogan performs excellent in image-level anomaly detection tasks, but does not perform as well as the other methods in pixel-level medical image detection tasks. AnoDDPM performs excellent pixel-level detection tasks, though sensitive to the selected thresholds. It is much slower than other methods in training and inference time. MS-FLOW further improves pixel-level image detection through spatial and channel attention mechanisms. Normalized flow-based characterization methods are not only fast to train but also yield satisfactory results even with small datasets.

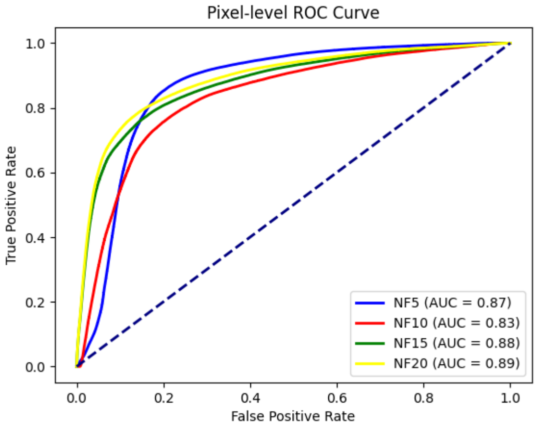

Fig. 4. a schematic diagram of the ROC curve on the NIBT data set

4.4. Ablution Study

Since the model consists of feature extraction and the NF module, we investigated the performance of the model after choosing different backbone networks and the structure of the NF module. For the backbone networks, we chose four backbone networks, including two deep convolutional networks, resnet18 and wideresnet50, and two visual transformer methods, cait and diet. For the evaluation criteria, we still use the five metrics, such as AUC, depicted in section 4.3. Table 3 shows the performance of the attentional flow when the sub-networks use different convolutions, and the backbone network is resnet18, wideresnet50, cait, and diet. Choosing a more complex convolutional neural network or employing a visual transformer does not improve performance.

The number of NF modules is also an essential hyperparameter for our method, which impacts the model's performance and efficiency. So we selected 4 different numbers of NF modules for the MS-FLOW, as a way to observe the effect it has on the model performance. Figure 4 shows the roc curves under different NF modules. We finally find that the model has the best overall performance when the number of NF modules equals 15. Figure 5 shows the Precision-Recall curves for different numbers of NF modules, and the best performance is also obtained when the number of NF modules is equal to 15.

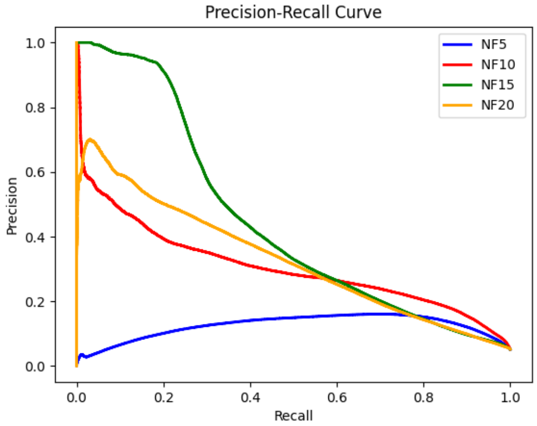

Fig. 5. a schematic diagram of the RP curve on the NIBT data set

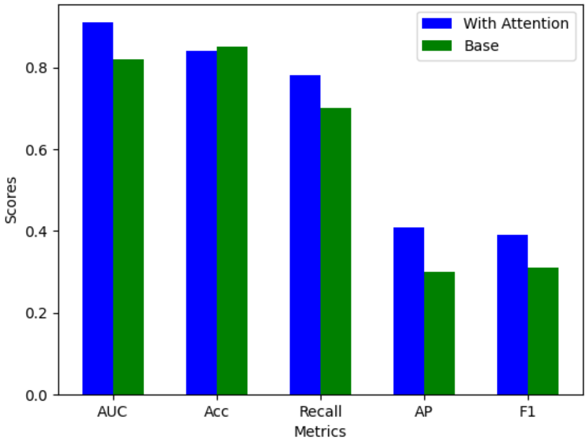

Fig. 6. Schematic representation of the attention mechanism for model performance images

To study the effectiveness of the proposed MS-NF module and observe the model performance changes after eliminating attention and adding attention mechanism in the sub-network. Figure 6 demonstrates the performance change of the model with and without the attention mechanism under the evaluation metrics such as AUC, and Acc. It can be seen that AUC, recall, AP, and F1 are improved, indicating that the attention module has a positive effect on the model.

4.5. Complexity evaluations

To analyze and evaluate the complexity of the model, we used the model parameter count and inference speed metrics to compare MS-NF with the other methods in Table 4. The comparison results are shown in Table N. MS-NF has a parameter count of 15.1M with the number of NF modules equal to 15, and a feature extraction network of resnet18, and an fps of 2.92 on the dataset. Compared to Anoddpm, MS-NF cuts the number of parameters to one-tenth of Anoddpm and improves the inference speed dramatically. Still, it is comparable to it in terms of detection effectiveness. In contrast, compared to fno-gan as well as fast-flow, the detection effect is better than the above two methods when the number of parameters is appropriately boosted.

It can be seen that the proposed MS-NF can achieve excellent performance and real-time detection with a small number of parameters, showing that the proposed work is competitive. The equipment configuration used for testing is composed of one Intel(R) Xeon(R) Silver 4314 CPU @ 2.40GHz, 128 GB memory, 4 TB storage, and one NVIDIA GeForce RTX 3090 card attached. All experiments are carried out in the operation system Linux. Pycharm is selected as the development environment. In addition, the program code of the models in this paper is implemented by Pytorch, an open-source tool that supports the agile development of deep learning models.

Table 4. Model Complexity Analysis

	parameters	time	fps
Ours	15.1M	9.26	2.92
f-AnoGan	3.93M	4.33	6.23
AnoDDPM	131M	0.04	632
fast-flow	15.0M	9.97	2.72

5. Concluding Remarks and Future Work

In this work, we propose a novel unsupervised learning model for medical image detection, which performs recognition without label information. It can avoid the subjective factors of the annotators to impact the performance of the model. Concretely, this model fuses the different scale feature maps to learn the contextual information and to detail multi-granularity information hidden in the data. Then, the normalizing flow in the model learns the distribution of the data, which is used for detection. Moreover, the convolutional attention mechanism makes the model focus on the useful regions. It can further improve the performance without significantly increasing the number of parameters. The experimental results illustrate that the proposed model outperforms the compared models. Particularly, the training speed of the proposed model is faster than others. The proposed model aims to learn the distribution of the normal medical images, which are used to generate new abnormal medical images combined with the abnormal information through the transfer module. It can relieve the requirement of the abnormal medical images. In future directions, we aim further to introduce the GAN mechanism to the proposed model to improve the quality of the generated abnormal medical images.

Acknowledgments. This work was partially supported by the National Key Research and Development Program of China under Grant 2021YFA1000600, the National Natural Science Foundation of China under Grant 62072170, the Science and Technology Project of Department of Communications of Hunan Provincial under Grant 202101, the Key Research and Development Program of Hunan Province under Grant 2022GK2015, and the Open Research Fund of Hunan Provincial Key Laboratory of Network Investigational Technology under grant 2020WLZC001.

References

1. Bhatt, N., Prados, D.R., Hodzic, N., Karanassios, C., Tizhoosh, H.R.: Unsupervised detection of lung nodules in chest radiography using generative adversarial networks. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 3842–3845. IEEE (2021)
2. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018)
3. Cai, J., Liang, W., Li, X., Li, K., Gui, Z., Khan, M.K.: Gtxchain: A secure iot smart blockchain architecture based on graph neural network. IEEE Internet of Things Journal 10(24), 21502 – 21514 (2023)
4. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A.: Generative adversarial networks: An overview. IEEE signal processing magazine 35(1), 53–65 (2018)
5. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real nvp. arXiv preprint arXiv:1605.08803 (2016)
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
7. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12873–12883 (2021)
8. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. Advances in neural information processing systems 30 (2017)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Ji, H., Xie, K., Wen, J., Zhang, Q., Xie, G., Liang, W.: Finemon: An innovative adaptive network telemetry scheme for fine-grained, multi-metric data monitoring with dynamic frequency adjustment and enhanced data recovery. Proc. ACM Manag. Data 2(1) (mar 2024), https://doi.org/10.1145/3639267
11. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4401–4410 (2019)

12. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
13. Li, C., He, A., Wen, Y., Liu, G., Chronopoulos, A.T.: Optimal trading mechanism based on differential privacy protection and stackelberg game in big data market. IEEE Transactions on Services Computing 16(5), 3550–3563 (2023)
14. Li, C., Peng, Y., Liu, G., Li, Y., Yang, X., Chen, C.: Efficient vision transformer for human-centric aiot applications through token tracking assignment. IEEE Transactions on Consumer Electronics 70(1), 1029–1039 (2024)
15. Li, C., Zhu, D., Hu, C., Li, X., Nan, S., Huang, H.: Ecdx: Energy consumption prediction model based on distance correlation and xgboost for edge data center. Information Sciences 643, 119218 (2023), https://www.sciencedirect.com/science/article/pii/S0020025523008034
16. Liang, W., Li, Y., Xie, K., Zhang, D., Li, K.C., Souri, A., Li, K.: Spatial-temporal aware inductive graph neural network for c-its data recovery. IEEE Transactions on Intelligent Transportation Systems 24(8), 8431 – 8442 (2022)
17. Liang, W., Li, Y., Xu, J., Qin, Z., Zhang, D., Li, K.C.: Qos prediction and adversarial attack protection for distributed services under dlaas. IEEE Transactions on Computers 73(3), 669 – 682 (2023)
18. Liang, W., Liu, Y., Yang, C., Xie, S., Li, K., Susilo, W.: On identity, transaction, and smart contract privacy on permissioned and permissionless blockchain: A comprehensive survey. ACM Comput. Surv. 56(12) (jul 2024), https://doi.org/10.1145/3676164
19. Liu, Y., Liang, W., Xie, K., Xie, S., Li, K., Meng, W.: Lightpay: A lightweight and secure off-chain multi-path payment scheme based on adapter signatures. IEEE Transactions on Services Computing 17(4), 17 (2023)
20. Long, J., Liang, W., Li, K.C., Wei, Y., Marino, M.D.: A regularized cross-layer ladder network for intrusion detection in industrial internet of things. IEEE Transactions on Industrial Informatics 19(2), 1747–1755 (2022)
21. Mishra, P., Verk, R., Fornasier, D., Piciarelli, C., Foresti, G.L.: Vt-adl: A vision transformer network for image anomaly detection and localization. In: 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE). pp. 01–06. IEEE (2021)
22. Pernet, C., Gorgolewski, K., Ian, W.: A neuroimaging dataset of brain tumour patients (2016)
23. Pinaya, W.H., Graham, M.S., Gray, R., Da Costa, P.F., Tudosiu, P.D., Wright, P., Mah, Y.H., MacKinnon, A.D., Teo, J.T., Jager, R., et al.: Fast unsupervised brain anomaly detection and segmentation with diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 705–714. Springer (2022)
24. Platt, J.: Sequential minimal optimization: A fast algorithm for training support vector machines (1998)
25. Puccio, B., Pooley, J.P., Pellman, J.S., Taverna, E.C., Craddock, R.C.: The preprocessed connectomes project repository of manually corrected skull-stripped t1-weighted anatomical mri data. Gigascience 5(1), s13742–016 (2016)
26. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
27. Razavi, A., Van den Oord, A., Vinyals, O.: Generating diverse high-fidelity images with vq-vae-2. Advances in neural information processing systems 32 (2019)
28. Rezende, D., Mohamed, S.: Variational inference with normalizing flows. In: International conference on machine learning. pp. 1530–1538. PMLR (2015)
29. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. arxiv 2021. arXiv preprint arXiv:2106.08265
30. Rudolph, M., Wandt, B., Rosenhahn, B.: Same same but differnet: Semi-supervised defect detection with normalizing flows. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1907–1916 (2021)

31. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. Medical image analysis 54, 30–44 (2019)
32. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International conference on information processing in medical imaging. pp. 146–157. Springer (2017)
33. Teng, Y., Li, H., Cai, F., Shao, M., Xia, S.: Unsupervised visual defect detection with score-based generative model. arXiv preprint arXiv:2211.16092 (2022)
34. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H.: Training data-efficient image transformers & distillation through attention. In: International conference on machine learning. pp. 10347–10357. PMLR (2021)
35. Van Den Oord, A., Vinyals, O., et al.: Neural discrete representation learning. Advances in neural information processing systems 30 (2017)
36. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems 30 (2017)
37. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
38. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 650–656 (2022)
39. Xie, S., Xiao, L., Han, D., Xie, K., Li, X., Liang, W.: Hcvc: A high-capacity off-chain virtual channel scheme based on bidirectional locking mechanism. IEEE Transactions on Network Science and Engineering 11(5), 3995 – 4006 (2023)
40. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. arXiv preprint arXiv:2111.07677 (2021)
41. Zhou, S., Li, K., Chen, Y., Yang, C., Liang, W., Zomaya, A.Y.: Trustbcfl: Mitigating data bias in iot through blockchain-enabled federated learning. IEEE Internet of Things Journal 11(15), 25648–25662 (2024)
42. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
43. Zimmerer, D., Kohl, S.A., Petersen, J., Isensee, F., Maier-Hein, K.H.: Context-encoding variational autoencoder for unsupervised anomaly detection. arXiv preprint arXiv:1812.05941 (2018)

Yufeng Xiao received the B.Sc. and M.S. degrees in computer application technology from Hunan University in 2010 and 2013, respectively, and the Ph.D. degree in computer science and technology from the same school in 2020. He is an assistant professor at the School of Computer Science and Engineering, Hunan University of Science and Technology. His current research interests include speech information processing, image processing and deep learning.

Xueting Huang is currently pursuing a master's degree at Hunan University of Science and Technology, with research focused on enhancing the overall performance of blockchain through sharding technology, as well as on information security and privacy protection in blockchain systems.

Wei Liang received a Ph.D. degree in computer science and technology from Hunan University in 2013. He is a Postdoctoral Scholar at Lehigh University from 2014 to 2016.

He is currently a Professor at the School of Computer Science and Engineering, Hunan University of Science and Technology. His research interests include intelligent transportation, security of IoV, blockchain, embedded systems and hardware IP protection, and security management in wireless sensor networks.

Jingnian Liu received an M.Sc. from Hunan University of Science and Technology. He has published papers (EI-indexed) and holds one invention patent. His research interests focus on computer vision and deep learning.

Yuxiang Chen received a PhD degree in computer application from Hunan University in 2021. He worked as a postdoctoral fellow in the Department of Computing at Hunan University from 2021.7 to 2023.3. He is currently an assistant professor at Hunan University of Science and Technology. His research interests include network monitoring, network security, big data, and AI.

Rui Xie received his Bachelor's and Master's degrees in 2006 and 2010, respectively, and is currently pursuing his Ph.D. at Hunan University of Science and Technology. His research interests include network security, cloud and edge computing, and AI, with a particular focus on distributed computing, privacy protection, and embedded systems.

Kuanching Li received a Ph.D. degree in 2001. He has authored or co-authored papers in high-ranked journals and conferences, and serves as Associate and Guest editor for several scientific journals. His research interests include cloud and edge computing, Big Data, and blockchain.

Nam Ling received a B.Eng. degree from the National University of Singapore in 1981 and the M.S. and Ph.D. degrees from the University of Louisiana at Lafayette, USA, in 1985 and 1989, respectively. He is currently the Associate Dean for Research for the School of Engineering and the Wilmot J. Nicholson Family Chair Professor at Santa Clara University (SCU), California, USA. His topics of interest include video coding algorithms and architectures.

Received: February 27, 2024; Accepted: December 01, 2024.