Journal of Network Intelligence ©2026 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **11**, Number **1**, February **2026**

Deep Image Steganalysis Network via Pulse-aware Excitation and Multi-level Feature Fusion for IIoT

Mengfei Chen[1,*], Zhenyu Liu[1], Zhenlin Zhai[1]

[1]School of Automation Engineering
Fujian College of Water Conservancy and Electric Power
Sanming 366000, China
1229754764@qq.com, 1434212422@qq.com, 842980717@qq.com

*Corresponding author: Mengfei Chen

Received May 16, 2025, revised September 03, 2025, accepted December 25, 2025.

Abstract. *With the widespread application of the industrial internet of things (IIoT) in fields such as smart manufacturing, which has greatly promoted industrial production, its information security issues have also gradually attracted extensive attention. Image steganalysis technology can detect whether digital images contain secret information, thereby effectively protecting the information security of the IIoT. In this paper, we propose a highly effective deep learning image steganalysis network, designated PMNet. In the preprocessing stage, to enhance the network's attention to steganographic signal-rich regions within feature maps while suppressing image content, and consequently improve detection accuracy, PMNet integrates a bottleneck residual block (BRB) with the attention mechanism module, pulse-aware excitation (PAE). This combination forms a pulse-aware excitation bottleneck residual block (PAEBEB), serving as the primary component of the preprocessing part. To further improve PMNet's detection accuracy, the network utilizes intermediate features from multiple feature*

extraction stages. These intermediate features are transformed into classification features through depthwise separable convolution (DWSConv) and global average pooling (GAP). These are then combined with classification features generated by multi-view global pooling (MGP) to constitute a multi-level feature fusion (MFF), which is subsequently fed into the classifier. Experimental results demonstrate that the proposed PMNet achieves detection accuracy on WOW, S-UNIWARD, and HILL steganography that is markedly superior to other comparative networks, reaching an accuracy as high as 92.52% on the WOW algorithm.

Keywords: Industrial Internet of Things, Steganalysis, Pulse-aware Excitation Bottleneck Residual Block, Multi-level Feature Fusion

1. **Introduction.** Industrial internet of things (IIoT) technology, an organic integration of internet of things technology and industrial automation technology, enables real-time monitoring and control of industrial equipment [1]. It is currently widely applied in scenarios such as smart grids [2], Intelligent Manufacturing [3], intelligent homes [4], and intelligent transportation [5] effectively enhancing enterprise production efficiency, improving product quality, and reducing production costs, thereby fostering socio-economic development and profoundly transforming human production and lifestyles. With the widespread adoption of IIoT technology, various devices within IIoT systems generate vast amounts of data. Some of this core data is critically important for enterprises, and

its leakage can lead to immeasurable losses. Consequently, the privacy protection of core data constitutes a major challenge confronting IIoT [6–8].

Steganography is a mature information protection technique that enables covert communication over public channels by embedding secret information into digital media [9]; critically, the stego media (digital media with embedded secret information) typically exhibits no perceptible visual changes. Digital images, characterized by high data redundancy and widespread availability on the internet, have become an ideal cover for steganography. Within the IIoT domain, image steganography can be employed to embed sensitive private information into digital images, thereby safeguarding data security [10, 11]. For example, Djebbar et al. [10] utilized image steganography to protect critical communication data within IIoT environments. However, technology is often a double-edged sword. With the continuous advancement of steganography, it is not only employed for legitimate information security purposes but has also inevitably become a tool for malicious actors [12]. Illicit individuals can exploit image steganography to exfiltrate various types of confidential data generated by IIoT systems during their operation. Moreover, the synergistic use of steganography to conceal malicious programs within digital images is emerging as a new trend in cyberattacks. The abuse of steganography poses a severe threat to the secure and stable operation of enterprises within IIoT application scenarios. Consequently, the detection of steganography in the IIoT domain to prevent data leakage and mitigate malicious attacks is of paramount importance.

Image steganalysis technology is used to determine whether digital images contain secret information, thereby detecting if digital images have been modified by steganography [13]. Therefore, image steganalysis technology can be utilized to supervise various digital images in IIoT systems, preventing digital images containing malicious programs from attacking IIoT systems, and simultaneously preventing relevant personnel from using steganography to steal various private data in IIoT systems. Consequently, researching image steganalysis technology is of great significance for preventing the abuse of image steganography and protecting the information security of IIoT systems.

Image steganalysis can be categorized into traditional steganalysis and deep learning-based steganalysis [12]. Traditional image steganalysis algorithms are mainly targeted at spatial-domain images [14–17] and joint photographic experts group (JPEG) images [18–20], typically rely on handcrafted high-dimensional features, and are divided into two parts: feature extraction and a classifier. With the continuous advancement of deep learning technology, deep learning-based image steganalysis has progressively become the mainstream approach in the field due to its superior detection performance, and can be divided into two types of methods: semi-learning [21–28] and full-learning [29–32]. Currently, in the preprocessing stage, deep learning-based steganalysis networks primarily employ spatial rich model (SRM) high-pass filters (HPF) or a combination of SRM and convolutional layers to suppress image content and enhance the signal-to-noise ratio (SNR) of the steganographic signal. These methods tend to give equal attention to feature maps from different channels, which can lead to an overemphasis on signals irrelevant to the steganalysis task, such as image content. In the feature extraction phase, convolution is used for feature extraction. Classification features are typically obtained from the last convolutional layer using global average pooling (GAP) [23, 24, 26, 30, 31] or specialized pooling techniques [25, 27, 28, 32]. This approach may overlook steganographic signal features extracted in the intermediate layers of the network or the types of steganographic signals acquired might be relatively uniform. To address these limitations and further improve accuracy, this paper introduces the PMNet steganalysis network. The main contributions are as follows:

(1) Pulse-aware excitation bottleneck residual block

A pulse-aware excitation bottleneck residual block (PAEBRB) is formed by combining the pulse-aware excitation (PAE) module [33] with the bottleneck residual block (BRB) [27]. This PAEBRB, along with the SRM, serves as the preprocessing part of PMNet. The PAEBRB can more effectively focus on feature maps of important channels and regions rich in steganographic signals within these feature maps. This allows for the extraction of more accurate steganographic signal features and reduces the influence of image content, thereby enhancing detection performance.

(2) Multi-level feature fusion

To further leverage and enrich the classification features fed into the classifier for improved detection performance, PMNet, in its feature extraction phase, processes intermediate features from multiple feature extraction stages. These intermediate features are first unified in dimension using depthwise separable convolution (DWSConv) and GAP. Subsequently, they are combined with the classification features generated by multi-view global pooling (MGP) [27] through a concat operation to form a multi-level feature fusion (MMF), which is then input to the classifier.

The remainder of this paper is organized as follows: Section 2 introduces the current research status of deep learning-based image steganalysis. Section 3 describes the network architecture and main components of the proposed PMNet. Section 4 presents a performance comparison of PMNet with other algorithms on various benchmark datasets. Finally, Section 5 provides a summary of the entire paper.

2. **Related work.** Although extensive early research led to high detection accuracy for traditional steganalysis, conventional image steganalysis relies on handcrafted high-dimensional features. This typically requires researchers to possess substantial prior knowledge. Furthermore, its feature extraction process and classifier training are conducted separately, meaning the classifier and the feature extraction process cannot mutually guide and optimize each other [21]. Thanks to the rapid development of deep learning, these shortcomings of traditional steganalysis have been effectively overcome by deep learning-based steganalysis methods. Currently, deep learning-based steganalysis methods are the mainstream approach in image steganalysis technology. Image steganalysis based on deep learning is generally divided into three stages: preprocessing, feature extraction, and classification. These three stages are uniformly optimized in an end-to-end manner, significantly reducing the reliance on researchers' prior knowledge and allowing researchers to focus on the design of the deep learning model architecture. At present, the performance of deep learning-based steganalysis methods has far surpassed that of traditional steganalysis methods.

Tan et al. [29] proposed Tan-Net, which marked the first attempt to introduce deep learning techniques into the field of image steganalysis. Tan-Net utilized three convolutional layers and used convolutional auto-encoder for pre-training. Although the performance of Tan-Net was lower than that of SRM, it pioneered new research techniques for image steganalysis. Qian et al. [21] introduced GNCNN, which used KV kernels in the preprocessing layer to enhance the SNR of the steganographic signal. Concurrently, considering the characteristics of steganographic signals, GNCNN incorporated average pooling and a Gaussian activation function. Experiments demonstrated that GNCNN achieved detection performance comparable to SRM. Xu et al. proposed XuNet [22]. XuNet employed KV kernels in the preprocessing stage and used an absolute value activation function to constrain the modeling range of the network. To prevent overfitting, a TanH activation function was used after the first convolutional layer, and 1×1 convolutional kernels were used in the last three convolutional layers.

Ye et al. [30] proposed a 10-layer steganalysis network called YeNet. This network utilizes 30 SRM HPF in its preprocessing part to suppress image content and propose the truncated linear unit (TLU) activation function to further enhance the SNR of the stego signal. YeNet also incorporated knowledge of selection channel, further improving detection accuracy. The detection performance of YeNet surpassed SRM by a significant margin. Yedroudj et al. [23] proposed Yedroudj-Net, which combined the advantages of XuNet and YeNet. It adopted SRM for preprocessing, and its feature extraction part consisted of 5 convolutional layers combined with TLU, rectified linear unit (ReLU), batch normalization (BN), and average pooling. The classifier part used a fully connected (FC) layer with a sufficiently large number of parameters. Experiments showed that Yedroudj-Net's detection performance comprehensively exceeded that of YeNet and XuNet. Li et al. [24] proposed ReTS-Net. ReTS-Net features three parallel SubNets with identical structures, but each SubNet is preprocessed using different HPF. Within the sub-networks, three activation functions—ReLU, Tanh, and Softmax—are used. Finally, the features extracted by the three SubNets are merged and fed into the classifier.

Boroumand et al. [31] proposed SRNet, which is composed of four different types of blocks, including residual blocks to prevent network degradation. SRNet does not use any HPF to process image content, and the parameters of all layers are randomly initialized before training. Consequently, SRNet can detect not only spatial domain images but also JPEG domain images. Zhu et al. [32] proposed ZhuNet. In ZhuNet, the SRM HPF in the preprocessing part were optimized and participated in the network updates. To improve detection accuracy, ZhuNet also utilized DWSConv and spatial pyramid pooling [34]. Deng et al. [25] proposed CovpoolNet. CovpoolNet features a carefully designed network architecture that significantly reduces training time. Furthermore, it employs global covariance pooling before classification to extract second-order statistic of the stego signal, thereby enhancing detection accuracy. You et al. [26] introduced SiaStegNet. SiaStegNet uses a siamese network architecture, dividing the image into two parts that are fed into two sub-networks with shared parameters. By analyzing the differences in the outputs of these two sub-networks, it determines whether the image is a stego image. This allows it to perform detection on images of arbitrary sizes.

Weng et al. [27] proposed a lightweight steganalysis network, LWENet. LWENet combines SRM HPF with BRB to enhance the SNR of the steganographic signal. To maintain its lightweight nature and improve accuracy, DWSConv is used in the final layer of the network, and MGP was introduced to enrich classification features. He et al. [28] proposed GFS-Net. GFS-Net strikes a good balance between performance, parameters, and training time. In the preprocessing stage, it employs gated channel transformation and pointwise convolution to boost the signal-to-noise ratio. The feature extraction part uses FasterNet blocks [35] to maintain a lightweight structure while improving accuracy. The classifier part utilizes Stylepooling to augment classification features, further enhancing performance.

3. The Proposed method.

3.1. PMNet.

To further enhance the detection accuracy of deep steganalysis networks, this paper proposes PMNet. Figure 1 illustrates the network architecture of PMNet. PMNet is composed of five different types of blocks. Block A, also known as PAEBRB, is a residual structure containing two convolutional layers with kernel sizes of 1×1 and 3×3, respectively. Both convolutional layers are preceded by BN and ReLU activation functions, and a PAE block follows the convolutional layers. Block B consists of a convolutional layer with a 1×1 kernel size, preceded by BN and ReLU. Block C is a convolutional layer with a 3 × 3 kernel size, followed by BN and ReLU. Block D is a DWSConv with a 3 × 3 kernel size, followed by BN and ReLU. Block E is a DWSConv with a 3 × 3 kernel size, followed by GAP.

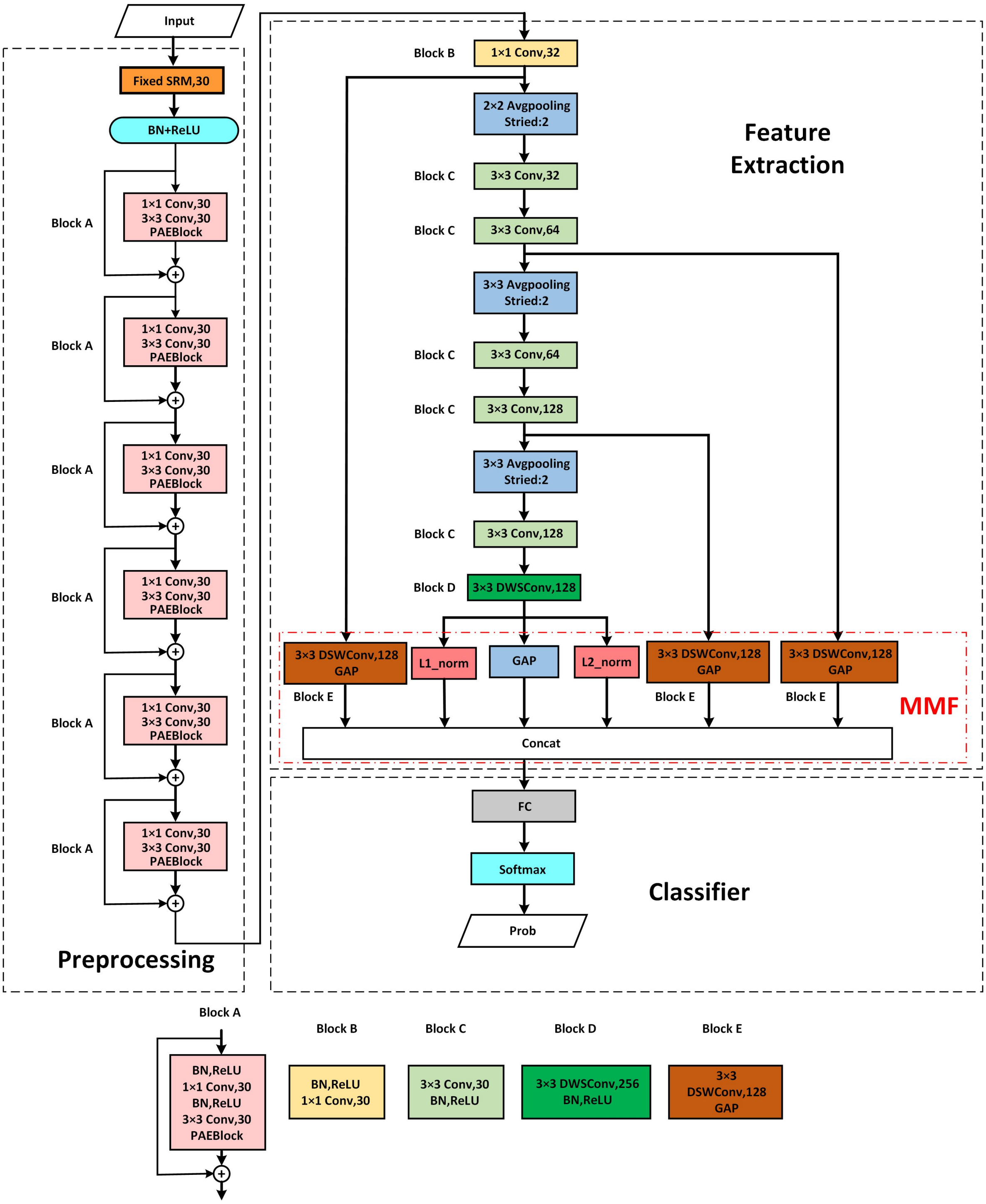


Figure 1. PMNet consists of three parts: preprocessing, feature extraction, and a classifier. $a \times a$ Conv, b denotes a convolutional layer with kernel size of $a \times a$ and b output feature maps. L1-norm, L2-norm, and GAP constitute MGP.

In the preprocessing part, the input image first passes through 30 SRM HPF to suppress image content and enhance the SNR of the stego signal. To maintain stability during network training, the SRM HPF are kept fixed and do not participate in updates during the training process. The features filtered by the SRM HPF, after passing through BN and ReLU, are fed into six PAEBRB to further extract features of the stego signal. To avoid weakening the faint stego signal, no pooling layers are used in the preprocessing part, and the number of channels in the convolutional layers is consistently maintained at 30 throughout this stage.

The feature extraction part is designed to extract high-level features of the stego signal. Initially, one Block B is employed to increase the number of feature channels to 32, followed by a 2×2 average pooling layer to reduce the feature map size and decrease the computational load. Subsequently, Block C is utilized to successively increase the feature channel count first to 64 and then to 128, accompanied by two 3×3 average pooling layers to further reduce the feature map dimensions. Following this, Block D raises the number of feature channels to 256. Ultimately, the MMF generates 1152-dimensional features through Block E and MGP, which are then fed into the classifier.

The classifier part is responsible for generating the classification result. PMNet utilizes a single FC layer followed by a Softmax function as its classifier. The classifier takes the 1152-dimensional classification features generated by MMF and converts them into 2-dimensional prediction probabilities. Finally, the classification result is output based on these prediction probabilities.

3.2. **PAEBRB.** To further suppress image content and enhance network detection accuracy, PMNet utilizes SRM+PAEBRB as its preprocessing part. While LWENet [27] effectively improved network detection accuracy by using SRM+BRB as its preprocessing section, a limitation exists: the 30 different kernels of the SRM HPF generate 30 feature map channels whose importance varies. Furthermore, the significance of pixels at different locations within each feature map is also inconsistent. Consequently, the standard BRB cannot effectively focus on the feature maps of important channels and the pixels at significant locations while suppressing useless information. To overcome the shortcomings of BRB, PMNet combines the attention mechanism PAE block [33] with BRB to form PAEBRB. Since the PAE block possesses both spatial attention and channel attention capabilities, PAEBRB can, through training, adaptively focus on the feature maps of important channels and the significant pixel locations within those feature maps. This enables the extraction of more accurate stego signal features, further improving the network's detection accuracy.

The structure of PAEBRB is shown as Block A in Figure 1 For an input feature map X_{input}, the output computation process of PAEBRB is as follows:

$$X_1 = conv_{1\times1}(relu(bn(X_{input}))) \tag{1}$$

$$X_2 = conv_{3\times3}(relu(bn(X_1))) \tag{2}$$

$$X_{output} = X_{input} + PAEBlock(X_2) \tag{3}$$

where $conv_{1\times1}$ and $conv_{3\times3}$ represent the convolutional layers within PAEBRB with kernel sizes of 1×1 and 3×3, respectively. X_1 and X_2 are the intermediate feature maps after passing through $conv_{1\times1}$ and $conv_{3\times3}$, respectively. $PAEBlock(\cdot)$ denotes the operation process of the PAE block, which will be detailed in Subsection 3.2.1. $relu(\cdot)$ denotes ReLU, $bn(\cdot)$ denotes BN. Activation X_{output} is the output feature map of the PAEBRB.

3.2.1. *PAE Block.* The structure diagram of the PAE block is shown in Figure 2. It comprises two parts: spatial attention and channel attention. After the input feature map passes through a convolution with a 1×1 kernel size and a Sigmoid function, the spatial attention map F_{S} is obtained, assigning larger weights to important regions within the feature map. After processing through GAP, two FC layers, and a Sigmoid function, the channel attention map F_{C} is obtained, assigning larger weights to important channels.

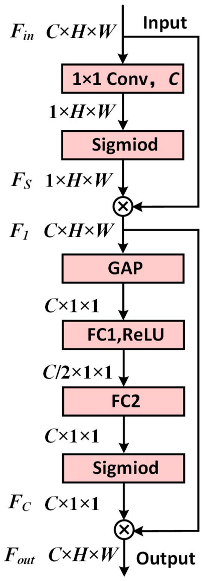


Figure 2. PAE Block. C, H, and W represent the number of channels, height, and width of the feature map, respectively.

For an input feature map F_{in} of size $C \times H \times W$, the computation process of the PAE block is as follows:

(1) Spatial Attention:

$$F_S = \sigma(conv_{1\times1}(F_{in})) \tag{4}$$

$$F_1 = F_S \otimes F_{in} \tag{5}$$

where $\sigma(\cdot)$ denotes the Sigmoid function, $\otimes$ represents the Hadamard product, $F_{in} \in \mathbb{R}^{C\times H\times W}$, $F_S \in \mathbb{R}^{1\times H\times W}$, $F_1 \in \mathbb{R}^{C\times H\times W}$.

(2) Channel Attention:

$$F_C = \sigma(fc_2(relu(fc_1(GAP(F_1))))) \tag{6}$$

$$F_{out} = F_C \otimes F_1 \tag{7}$$

where fc denotes a fully connected layer, fc_1 has an input dimension of C and an output dimension of $C/2$, fc_2 has an input dimension of $C/2$ and an output dimension of C, and $GAP(\cdot)$ represents the global average pooling operation, $F_C \in \mathbb{R}^{C\times1\times1}$, $F_{out} \in \mathbb{R}^{C\times H\times W}$.

3.3. **MMF.** Most steganalysis networks utilize GAP or specialized types of pooling before the classifier to compress the features extracted from the final layer of the feature extraction part, thereby obtaining classification features. However, within the feature maps generated during feature extraction, those from different levels possess information with distinct characteristics, and their contributions to the classification task also vary [36]. Utilizing feature maps from different levels such that they complement each other can significantly enrich the classification features fed into the classifier, thus enhancing detection accuracy. Obtaining classification features solely from the final layer might overlook the valuable contributions of intermediate layer features to the classification, or it could result in the acquisition of features that are relatively uniform in type.

This paper proposes MMF to capture classification features from different network levels; the structure of MMF is illustrated within the red dashed box in Figure 1 MMF utilizes multiple intermediate features from the feature extraction part, ensuring information richness by sourcing these features before any average pooling operations. Through DWSConv with a 3×3 kernel size, the channel dimensions of these multiple intermediate features are unified to 128. Subsequently, GAP is applied (as defined in Block E) to obtain classification features from each intermediate path. The use of DWSConv significantly reduces the number of additional parameters introduced by incorporating these intermediate features. Considering that MGP has demonstrated strong performance in the steganalysis domain, MMF also integrates MGP applied to the output features of Block D. MGP [27] includes L1-norm pooling, L2-norm pooling, and GAP, compressing the 256-channel output of Block D from multiple perspectives to generate 768-dimensional (=256×3) classification features. Finally, MMF employs a concatenation operation to combine the features derived from the three intermediate paths (each processed via a structure like Block E, resulting in 128×3=384 dimensions) and the features generated by MGP (768 dimensions). This results in a total of 1152 (=384+768) dimensional features, which are fed into the classifier. This approach substantially enriches the classification features and enhances detection accuracy.

4. **Experiment.**

4.1. **Datasets and Hyperparameters.**

To evaluate the performance of PMNet, we employ two standard datasets, BOSSBase [37] and BOWS2 [38], along with three spatial domain steganographic algorithms: WOW [39], S-UNIWARD [40], and HILL [41]. Each dataset initially contains 10,000 grayscale images in PGM format with size 512×512. Consistent with mainstream practices, we resize all images to 256×256 using MATLAB's imresize() function to serve as the cover images. Subsequently, for each cover image source (BOSSBase and BOWS2), corresponding stego images are generated using the three algorithms (WOW, S-UNIWARD, HILL) at four different embedding rates: 0.1, 0.2, 0.3, and 0.4 bits per pixel (bpp). This process results in 12 distinct experimental settings.

The 10,000 cover-stego pairs originating from BOSSBase are split into training, validation, and testing sets using a 4:1:5 ratio (resulting in 4,000 training pairs, 1,000 validation pairs, and 5,000 testing pairs). The 10,000 cover-stego pairs originating from BOWS2 are allocated entirely to the training set. Therefore, for each specific algorithm and embedding rate combination, the final dataset composition is as follows:

Training Set: 14,000 cover-stego pairs (4,000 from BOSSBase + 10,000 from BOWS2). Validation Set: 1,000 cover-stego pairs (from BOSSBase). Testing Set: 5,000 cover-stego pairs (from BOSSBase).

During the training process, data augmentation is applied to the training set. This augmentation includes random rotations (0°, 90°, 180°, 270°) and random horizontal flipping. Generally, PMNet is trained for a total of 200 epochs with a batch size of 32 (composed of 16 cover-stego pairs). The Stochastic Gradient Descent (SGD) optimizer is employed with an initial learning rate of 0.01, a momentum of 0.9, and a weight decay of 0.0005. The learning rate is scheduled to decay to 10% of its preceding value at epochs 81, 111, and 181. L2 regularization is disabled for the biases in all convolutional and FC layers. It

Table 1. Detection Accuracy Comparison: PMNet, SRNet, CovpoolNet, and LWENet. Bold black font denotes the best accuracy for the respective embedding rate.

Network	WOW				S-UNIWARD				HILL			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
SRNet	75.60	84.58	88.77	91.22	69.45	80.70	86.27	90.09	67.89	76.62	82.59	85.96
CovpoolNet	74.29	83.16	88.03	91.24	68.94	80.02	86.27	90.24	67.55	76.53	82.34	85.38
LWENet	76.23	85.23	89.63	92.15	69.27	80.93	86.58	90.29	69.04	77.49	82.96	86.57
PMNet	**76.69**	**85.85**	**89.89**	**92.52**	**71.10**	**81.51**	**87.18**	**91.05**	**69.46**	**78.21**	**83.62**	**87.01**

Table 2. AUC value Comparison: PMNet, SRNet, CovpoolNet, and LWENet. Bold black font denotes the best value for the respective embedding rate.

Network	WOW				S-UNIWARD				HILL			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
SRNet	85.79	93.75	96.60	97.91	77.56	90.17	94.88	97.34	75.84	86.24	91.63	94.62
CovpoolNet	84.85	93.14	96.36	98.01	78.62	90.36	95.39	97.58	76.94	87.28	92.40	95.08
LWENet	86.90	94.55	97.30	98.38	79.17	91.41	95.67	97.62	78.93	88.34	92.95	95.51
PMNet	**87.73**	**95.17**	**97.56**	**98.56**	**80.17**	**91.71**	**95.94**	**98.01**	**79.47**	**88.74**	**93.65**	**95.87**

should be noted that for training PMNet on the S-UNIWARD and HILL specifically at an embedding rate of 0.1 bpp, a modified approach was used: the model was initialized with network parameters from its training at 0.2 bpp and was subsequently trained for only 100 epochs. For these specific 0.1 bpp cases, the initial learning rate was set to 0.005, and this learning rate decayed to 10% of its preceding value at the 41st and 81st epochs. Other hyperparameters, such as the optimizer type, weight decay, and momentum, remained unchanged from the general settings. All experiments reported in this paper were conducted using an Nvidia RTX 3090 GPU and an Intel Core i5-14600KF CPU. The network implementations are based on PyTorch version 1.10.0, using Python version 3.8.0.

4.2. **Performance Comparison.** To evaluate the performance of PMNet, we compared its detection accuracy, receiver operating characteristic (ROC) curves, and area under the ROC curve (AUC) values with several state-of-the-art steganalysis networks: LWENet, SRNet, and CovpoolNet. To ensure a fair comparison, all networks were trained and tested under identical experimental conditions.

Table 1 presents a comparison of detection accuracy between PMNet and LWENet, SRNet, and CovpoolNet. As can be seen from Table 1, PMNet achieves the best detection accuracy compared to the other three networks across all tested embedding rates for the three steganographic algorithms. For example, on the S-UNIWARD 0.4 bpp dataset, PMNet's detection accuracy is 91.05%, which is 0.76%, 0.81%, and 0.96% higher than that of LWENet, SRNet, and CovpoolNet, respectively.

In Figure 3 and Table 2, we compare the ROC curves and AUC values of PMNet with those of LWENet, SRNet, and CovpoolNet. As can be seen from Figure 3, PMNet's ROC curves are closer to the top-left corner than those of the other three networks across various embedding rates for the three steganographic algorithms. Table 2 shows that PMNet also achieves the highest AUC values. The results presented in Table 1, Table 2, and Figure 3 collectively demonstrate that PMNet exhibits superior performance compared to the other networks.

4.3. **Performance under Algorithmic Mismatch.** In steganalysis tasks, algorithm mismatch scenarios frequently occur; for instance, a network trained on data from the

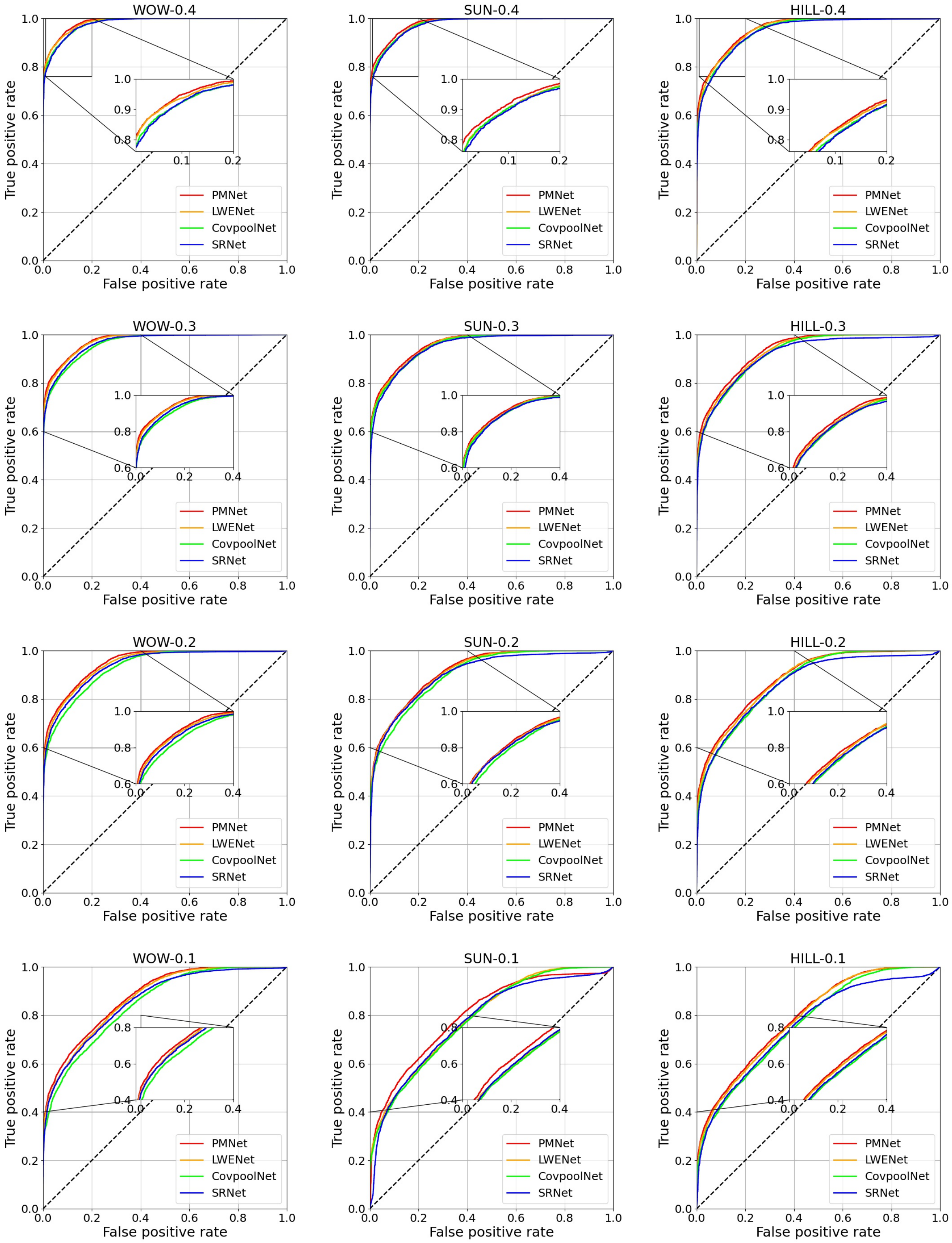


Figure 3. Comparison of ROC curves for PMNet, SRNet, CovpoolNet, and LWENet. For instance, "WOW-0.4" denotes the comparison of these four networks using the WOW algorithm at an embedding rate of 0.4 bpp.

Table 3. PMNet's Detection Performance under Algorithm Mismatch at 0.4 bpp

Train	Test		
	WOW	S-UNIWARD	HILL
WOW	92.52	80.44	63.87
S-UNIWARD	88.52	91.05	70.87
HILL	85.20	77.10	87.01

Table 4. PMNet's Detection Performance under Algorithm Mismatch at 0.2 bpp

Train	Test		
	WOW	S-UNIWARD	HILL
WOW	85.85	70.23	59.41
S-UNIWARD	81.26	84.51	65.95
HILL	74.09	65.12	78.21

Table 5. Performance comparison of PMNet when using BRB and PAEBRB, respectively; Steganographic algorithm: WOW.

Network	0.2bpp	0.4bpp
PMNet with BRB	85.39	92.38
PMNet with PAEBRB	85.85	92.52

WOW might be used to detect images embedded using the HILL. In Table 3 and Table 4, we evaluate PMNet's detection performance under such algorithm mismatch conditions, specifically at embedding rates of 0.4 bpp and 0.2 bpp, respectively. It can be observed that when faced with algorithm mismatch, PMNet's detection performance inevitably decreases. For example, when trained with the WOW at 0.4 bpp, its accuracy for detecting steganography by the S-UNIWARD is 80.44%. Nevertheless, PMNet still maintains a notable level of detection capability even under mismatched conditions, which indicates that PMNet possesses a degree of generalization ability for steganalysis tasks.

4.4. **Ablation Study.**

4.4.1. *Effectiveness of PAEBRB.* To evaluate the effectiveness of PAEBRB, we compare in Table 5 the detection performance of PMNet when using PAEBRB versus BRB [27]. It should be noted that BRB is essentially PAEBRB with the PAE block [33] removed. The results in Table 5 indicate that PMNet using PAEBRB achieves performance improvements of 0.46% and 0.14% at 0.2 bpp and 0.4 bpp, respectively, compared to when using BRB. This clearly demonstrates that PAEBRB effectively enhances PMNet's detection performance by focusing on the feature maps of important channels and on regions within these feature maps that are rich in stego signals.

4.4.2. *Effectiveness of MMF.* To demonstrate the effectiveness of MMF, in the experiments presented in Table 6, we utilized different methods to enrich classification features. 'PMNet with MMF' indicates that PMNet uses MMF to enrich classification features. 'PMNet with GAP and Block E' signifies that PMNet enriches classification features using GAP in conjunction with features from Block E paths. 'PMNet with MGP' indicates that PMNet uses MGP to enrich classification features. Compared to 'PMNet with GAP and Block E', MMF improves detection accuracy by 0.47% and 0.29% at 0.2 bpp and 0.4 bpp, respectively. Compared to 'PMNet with MGP', MMF enhances detection accuracy by 0.06% and 0.39% at 0.2 bpp and 0.4 bpp, respectively. These results fully demonstrate that using MMF in PMNet to enrich classification features effectively improves detection performance compared to the other evaluated methods.

Table 6. Performance comparison of PMNet using different methods to enrich classification features; Steganographic algorithm: WOW.

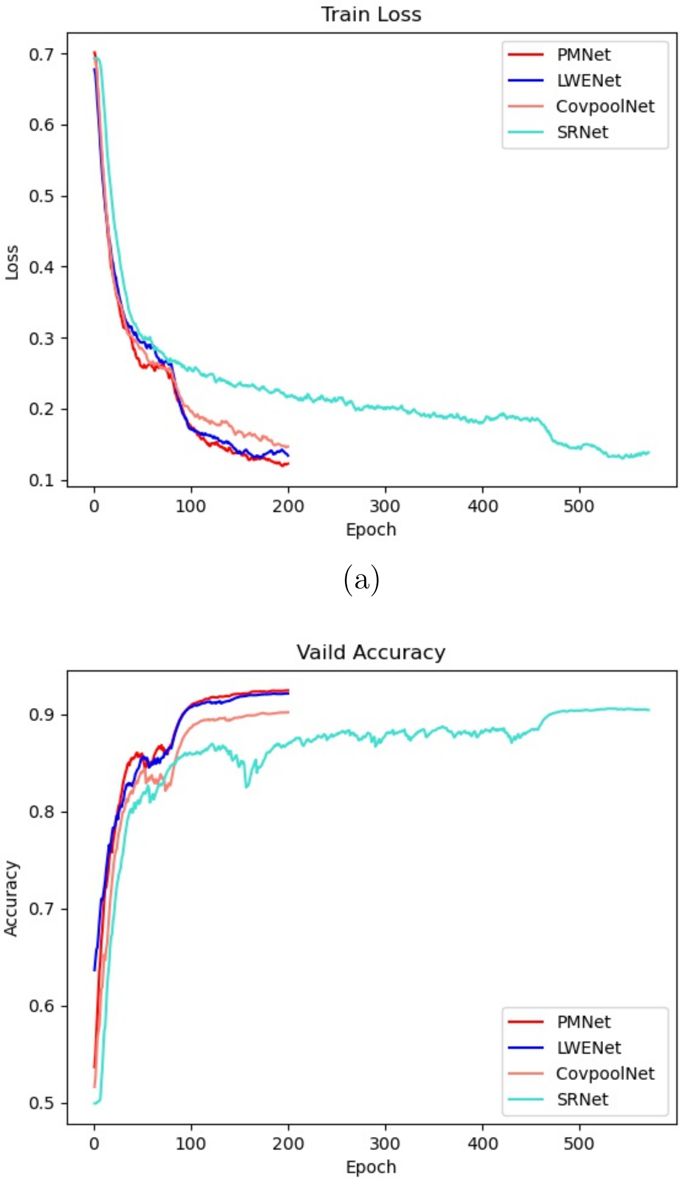
Network	0.2bpp	0.4bpp
PMNet with MMF	85.85	92.52
PMNet with GPA and Block E	85.38	92.23
PMNet with MGP	85.79	92.13

4.5. **Convergence Performance and the Number of Parameters Comparison.** Figure 4 shows the training loss curves and validation set accuracy curves for PMNet, LWENet, SRNet, and CovpoolNet when using the WOW steganographic algorithm at an embedding rate of 0.4 bpp. From Figure 4(a), it can be observed that all four networks converge well during the training process, but PMNet achieves a lower final loss value. Figure 4(b) indicates that PMNet's accuracy on the validation set is also the highest after multiple epochs. From this, it can be concluded that PMNet holds distinct advantages over the other three networks. Table 7 presents the number of parameters and the average test times on the test set for these networks. Notably, both PWNet and LWENet feature a relatively small and comparable number of parameters, which is advantageous for their deployment on IIoT devices. Although PMNet has the longest test time, it achieves the best detection performance.

Table 7. Comparison of the number of parameters and test time among PMNet, SRNet, CovpoolNet, and LWENet

Network	Number of parameters(M)	Averaged Test time(s)
SRNet	4.78	18.25
CovpoolNet	0.69	12.07
LWENet	0.38	18.62
PMNet	0.42	25.48

5. **Conclusion.** During the operation of IIoT systems, vast amounts of data are generated. To protect core data and prevent the misuse of steganography within IIoT systems, this paper focuses on research into image steganalysis technology and proposes PMNet. In this work, we combine the attention mechanism module, PAE, with BRB to form PAEBRB, which is utilized in the preprocessing stage of PMNet. This approach allows for focused attention on the feature maps of important channels and on regions within these feature maps that are rich in steganographic signals, thereby effectively enhancing the network's detection accuracy. Furthermore, this paper introduces MMF, which uses multiple intermediate features from the feature extraction stages in conjunction with MGP to enrich the classification features. This further boosts PMNet's detection performance. Experiments demonstrate that the proposed PMNet achieves strong detection performance across various embedding rates for three different steganographic algorithms, significantly outperforming other compared networks.



(a)

(b)

Figure 4. (a) Training loss curves for the four networks. (b) Validation accuracy curves for the four networks.

Acknowledgment. This work is supported by the Fujian College of Water Conservancy and Electric Power – College-level Fund: General Project (Category C). (No. YJKJ2426C) and Fujian Provincial Educational Research Project for Young and Middle-Aged Teachers(No. JAT231218).

REFERENCES

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.

[2] S. Al-Rubaye, E. Kadhum, Q. Ni, and A. Anpalagan, "Industrial internet of things driven by sdn platform for smart grid resiliency," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 267–277, 2019.

[3] Y. Hu, Q. Jia, Y. Yao, Y. Lee, M. Lee, C. Wang, X. Zhou, R. Xie, and F. R. Yu, "Industrial internet of things intelligence empowering smart manufacturing: A literature review," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19 143–19 167, 2024.

[4] Z. Liu, "Research on intelligent home care system based on internet of things," in *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, 2022, pp. 326–328.

[5] T.-Y. Wu, H. Wu, M. Tang, S. Kumari, and C.-M. Chen, "Unleashing the potential of metaverse in social iov: An authentication protocol based on blockchain," *CMC-Computers, Materials & Continua*, vol. 84, no. 2, pp. 3175–3192, 2025.

[6] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4177–4186, 2020.

[7] X. Zhang, X. Chen, J. K. Liu, and Y. Xiang, "Deeppar and deepdpa: Privacy preserving and asynchronous deep learning for industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2081–2090, 2020.

[8] T.-Y. Wu, H. Wu, S. Kumari, and C.-M. Chen, "An enhanced three-factor based authentication and key agreement protocol using puf in iomt," *Peer-to-Peer Networking and Applications*, vol. 18, no. 2, p. 83, 2025.

[9] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "Digital image steganography: Survey and analysis of current methods," *Signal Processing*, vol. 90, no. 3, pp. 727–752, 2010.

[10] M. Hassaballah, M. A. Hameed, A. I. Awad, and K. Muhammad, "A novel image steganography method for industrial internet of things security," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7743–7751, 2021.

[11] F. Djebbar, "Securing iot data using steganography: A practical implementation approach," *Electronics*, vol. 10, no. 21, 2021.

[12] J. Chen, Z. Fu, W. Zhang, X. Cheng, and X. Sun, "Review of image steganalysis based on deep learning," *Journal of Software*, vol. 32, no. 2, pp. 551–578, 2021.

[13] W. M. Eid, S. S. Alotaibi, H. M. Alqahtani, and S. Q. Saleh, "Digital image steganalysis: Current methodologies and future challenges," *IEEE Access*, vol. 10, pp. 92 321–92 336, 2022.

[14] T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," in *Proceedings of the 11th ACM Workshop on Multimedia and Security*, 2009, p. 75–84.

[15] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

[16] V. Holub and J. Fridrich, "Random projections of residuals for digital image steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 1996–2006, 2013.

[17] W. Tang, H. Li, W. Luo, and J. Huang, "Adaptive steganalysis against wow embedding algorithm," in *Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security*, 2014, p. 91–96.

[18] V. Holub and J. Fridrich, "Low-complexity features for jpeg steganalysis using undecimated dct," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, 2015.

[19] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, "Steganalysis of adaptive jpeg steganography using 2d gabor filters," in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, 2015, p. 15–23.

[20] T. D. Denemark, M. Boroumand, and J. Fridrich, "Steganalysis features for content-adaptive jpeg steganography," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1736–1746, 2016.

[21] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Media Watermarking, Security, and Forensics 2015*, vol. 9409. SPIE, 2015, pp. 171–180.

[22] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 708–712, 2016.

[23] M. Yedroudj, F. Comby, and M. Chaumont, "Yedroudj-net: An efficient cnn for spatial steganalysis," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2092–2096.

[24] B. Li, W. Wei, A. Ferreira, and S. Tan, "Rest-net: Diverse activation modules and parallel subnets-based cnn for spatial image steganalysis," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 650–654, 2018.

[25] X. Deng, B. Chen, W. Luo, and D. Luo, "Fast and effective global covariance pooling network for image steganalysis," in *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, 2019, p. 230–234.

[26] W. You, H. Zhang, and X. Zhao, "A siamese cnn for image steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 291–306, 2021.

[27] S. Weng, M. Chen, L. Yu, and S. Sun, "Lightweight and effective deep image steganalysis network," *IEEE Signal Processing Letters*, vol. 29, pp. 1888–1892, 2022.

[28] J. He, S. Weng, L. Yu, C. Zhang, and W. Chen, "An image steganoganalyzer with comprehensive detection performance," *IEEE Signal Processing Letters*, vol. 30, pp. 1682–1686, 2023.

[29] S. Tan and B. Li, "Stacked convolutional auto-encoders for steganalysis of digital images," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 2014, pp. 1–4.

[30] J. Ye, J. Ni, and Y. Yi, "Deep learning hierarchical representations for image steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2545–2557, 2017.

[31] M. Boroumand, M. Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2019.

[32] R. Zhang, F. Zhu, J. Liu, and G. Liu, "Depth-wise separable convolutions and multi-level pooling for an efficient spatial cnn-based steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1138–1150, 2020.

[33] Y. Qiu, H. Tian, H. Li, C.-C. Chang, and A. V. Vasilakos, "Separable convolution network with dual-stream pyramid enhanced strategy for speech steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2737–2750, 2023.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[35] J. Chen, S.-h. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H. G. Chan, "Run, don't walk: Chasing higher flops for faster neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 12 021–12 031.

[36] C. Xu, Y. Wang, W. Wang, G. Li, Y. Zheng, and Q. Zhang, "Multi-level features weighted fusion algorithm for domestic garbage image classification," *Journal of Chongqing University of Technology(Natural Seience)*, vol. 36, no. 9, pp. 146–155, 2022.

[37] P. Bas, T. Filler, e. T. Pevný, Tomáš", T. Pevný, S. Craver, and A. Ker, ""break our steganographic system": The ins and outs of organizing boss," in *Information Hiding*, 2011, pp. 59–70.

[38] P. Bas and T. Furon, "Bows-2 contest (break our watermarking system)," http://bows2.ec-lille.fr, 2008.

[39] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2012, pp. 234–239.

[40] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–13, 2014.

[41] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 4206–4210.