Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.067252

ARTICLE

Image Steganalysis Based on an Adaptive Attention Mechanism and Lightweight DenseNet

Zhenxiang He*, Rulin Wu and Xinyuan Wang

School of Cyberspace Security, Gansu University of Political Science and Law, Lanzhou, 730070, China

*Corresponding Author: Zhenxiang He. Email: hzx6198@gsupl.edu.cn

Received: 28 April 2025; Accepted: 12 June 2025; Published: 29 August 2025

ABSTRACT: With the continuous advancement of steganographic techniques, the task of image steganalysis has become increasingly challenging, posing significant obstacles to the fields of information security and digital forensics. Although existing deep learning methods have achieved certain progress in steganography detection, they still encounter several difficulties in real-world applications. Specifically, current methods often struggle to accurately focus on steganography sensitive regions, leading to limited detection accuracy. Moreover, feature information is frequently lost during transmission, which further reduces the model's generalization ability. These issues not only compromise the reliability of steganography detection but also hinder its applicability in complex scenarios. To address these challenges, this paper proposes a novel deep image steganalysis network designed to enhance detection accuracy and improve the retention of steganographic information through multilevel feature optimization and global perceptual modeling. The network consists of three core modules: the preprocessing module, the feature extraction module, and the classification module. In the preprocessing stage, a Spatial Rich Model (SRM) filter is introduced to extract the high-frequency residual information of the image to initially enhance the steganographic features; at the same time, a lightweight Densely Connected Convolutional Networks (DenseNet) structure is proposed to enhance the effective transmission and retention of the features and alleviate the information loss problem in the deep network. In the feature extraction stage, a hybrid modeling structure combining depth-separated convolution and ordinary convolution is constructed to improve the feature extraction efficiency and feature description capability; in addition, a dual-domain adaptive attention mechanism integrating channel and spatial dimensions is designed to dynamically allocate feature weights to achieve precise focusing on the steganography-sensitive region. Finally, the classification module adopts dual fully connected layers to realize the effective differentiation between coverage and steganography maps. These innovative designs not only effectively improve the accuracy and generalization ability of steganography detection, but also provide a new efficient network structure for the field of steganalysis. Numerous experimental results show that the detection performance of the proposed method outperforms the existing mainstream methods, such as SR-Net, TSNet, and CVTStego-Net, on the publicly available dataset BOSSbase and BOSW2. Meanwhile, multiple ablation experiments further validate the validity and reasonableness of the proposed network structure. These results not only promote the development of steganalysis technology but also provide more reliable detection tools for the fields of information security and digital forensics.

KEYWORDS: Image steganalysis; lightweight densenet; adaptive attention; feature focusing; information retention

1 Introduction

Image steganography is an information-hiding technique that embeds secret data into carrier images and is now widely applied to various multimedia formats such as images, audio, and text. Unlike traditional

encryption methods, image steganography aims to make the embedded information visually imperceptible, thereby enabling covert communication [1]. However, the misuse of steganography poses a serious threat to information security and personal privacy. According to [2], Cybercriminals often exploit steganography to evade regulations; even when the relevant evidence is encrypted, it remains difficult to detect. Therefore, as highlighted in [3], conducting in-depth research on effective image steganalysis and corresponding defense strategies is of great importance for safeguarding cybersecurity.

To perform steganalysis, it is generally necessary to select and extract features from both cover and stego images [4]. Traditional image steganalysis techniques [5–8] rely on pixel correlations, frequency domain coefficient distributions, and other statistical patterns. These methods manually extract features and analyze image statistics to build classifiers, primarily depending on hand-crafted features. Although such methods can effectively detect steganographic content to some extent, they exhibit notable limitations in feature design, generalization ability, and adaptability to complex scenes.

In recent years, deep learning-based approaches [9–11] have gradually become mainstream. These methods automatically learn potential steganographic traces (e.g., high-frequency noise distributions and texture perturbation patterns) through deep neural networks. They not only reduce the burden of manual feature design but also significantly improve the detection accuracy. However, deep learning methods still face challenges in real-world applications. It is hard to focus on small steganographic changes in images, and important information is often lost during training, which leads to less effective detection.

To address these challenges, "A novel steganalysis framework is proposed, designed to enhance the network's ability to focus on steganographic regions and preserve steganographic information without significantly increasing the number of network parameters". Specifically, A lightweight DenseNet is employed to prevent the loss of critical information in deep networks, and an adaptive attention mechanism is designed that fuses channel and spatial domain features to dynamically adjust feature weights. The overall network architecture enables improved focus on steganographically sensitive regions, mitigates the loss of steganographic information, and enhances both detection accuracy and generalization capability. This work provides new insights into efficient and accurate image steganalysis.

The main contributions of our work are as follows:

- A novel steganalysis network is introduced that effectively mitigates steganographic information loss and enhances the detection capability for low embedding rate steganography through multilevel feature optimization and global perceptual modeling.
- The proposed Lightweight DenseNet, named "DenseLite", is designed to retain more features and reduce information loss during feature propagation. This design significantly improves steganalysis detection accuracy by ensuring that more information is preserved in the preprocessing stage.
- An adaptive attention mechanism is designed to accurately focus on steganography-sensitive regions. By integrating channel and spatial attention, it dynamically learns to assign appropriate weights to each branch, thereby enhancing the model's ability to capture steganographic information and improve generalization.

The subsequent sections of this paper are structured as follows: Section 2, 'Related Work', reviews the relevant classical approaches. Section 3, 'Methods', describes our step analysis scheme. Section 4, 'Experimental Results', presents the experimental setup, along with the results and analysis. Section 5, 'Conclusions', summarizes the findings of this paper.

2 Related Work

2.1 Traditional Methods of Steganalysis

To extract steganographic information from images, some traditional approaches have been proposed. For example, Fridrich and Kodovsky [12] introduced the null domain-rich model, which obtains the residual information of an image by exploiting both linear and nonlinear relationships between the center pixel and its surrounding pixels. Denemark et al. [9] proposed a variant of the popular spatial richness model (SRM) for the selection channel, constructing joint high-order statistics of residuals from neighboring noise as a statistical descriptor to improve the detection of steganographic images. Ma et al. [10] proposed a feature selection method for steganalysis based on multidimensional evaluation and dynamic threshold assignment, which not only achieves high detection accuracy and effective feature dimensionality reduction but also eliminates dependence on specific classifiers.

However, as noted in, a major challenge of traditional methods lies in effectively extracting features relevant to steganographic content. Although these approaches have achieved certain success, their feature extraction processes can not be optimized through end-to-end training, which limits the model's learning capability. Moreover, the extracted information is often inadequate when dealing with complex and evolving steganographic algorithms. These limitations have significantly accelerated the development of deep learning-based approaches in the field of steganalysis.

2.2 Deep Learning-Based Steganalysis

There have been many studies related to image steganalysis based on deep learning. Xu et al. [11] proposed Xu-Net, which enhances statistical modeling in the subsequent layers by applying absolute value operations to the feature maps generated in the first convolutional layer. To prevent overfitting, they utilized a saturated region of the hyperbolic tangent (TanH) function in the early stages of the network to further constrain the range of data values. Additionally, 1×1 convolutions were employed in deeper layers to reduce modeling complexity, resulting in significantly improved detection accuracy compared to the SRM. Ye et al. [13] introduced Ye-Net, which used the 30 high-pass filter kernels from the SRM as a preprocessing layer. The network also incorporated truncated linear units and, for the first time, combined the selective channel technique with a neural network. The model demonstrated superior performance in detecting adaptive steganography compared to traditional SRM methods. Boroumand et al. [14] proposed SRNet, an end-to-end deep learning model that integrates residual modules into the network, further improving steganography detection accuracy. That same year, Yang et al. [15] designed a 32-layer convolutional neural network (CNN) similar to DenseNet [16], which improved preprocessing efficiency and feature reuse by connecting feature maps of the same size across layers. Zhang et al. [17] proposed Zhu-Net, which employed depthwise separable convolutions to extract both null domain and channel residual information. It also introduced the pyramid pooling module [18], enabling multi-scale pooling operations to capture more representative information. Although these methods have notably improved the extraction of steganographic features, they still face challenges in preserving weak steganographic signals during training.

In recent years, attention mechanisms have been increasingly introduced into image analysis tasks to enhance the model's ability to focus on critical information. By leveraging global statistical correlations to mine locally sensitive features from steganographic noise, attention-based techniques have advanced the development of image analysis. Weng et al. [19] propose a new deep steganalysis network called SwT-SN. It uses Directional Differential Adaptive Combining (DDAC) and three residual blocks to improve feature extraction. The network also includes a size-independent detector (CSPP-SID) and a two-part Swing Transformer (SwT) for better steganalysis. The goal is to enhance detection accuracy for steganographic

images of any size while reducing computational costs in training and testing. He et al. [20] proposed a unified steganalysis network called ESNet, which aims to effectively preserve and identify steganalysis signals from both the spatial and JPEG domains. Specifically, the network adopts a dual-branch preprocessing architecture: branch 1 uses a fixed SRM (Steganalysis Rich Model) kernel to extract noise residuals, while branch 2 uses a randomly initialized kernel to extract noise residuals. After that, the network will fuse the features of the two branches and exchange the fused complementary information through two carefully designed bidirectional fusion blocks. This process can effectively improve the signal-to-noise ratio, thereby better identifying steganalysis signals. Vaswani et al. [21] proposed the Transformer, introducing the self-attention mechanism, which completely discards traditional convolution and recurrent structures and relies solely on self-attention. The model achieved state-of-the-art performance in machine translation and marked the first successful application of self-attention in sequence modeling. The Convolutional Block Attention Module (CBAM) proposed by Woo et al. [22] sequentially integrates channel and spatial attention mechanisms. It effectively guides the network to focus on salient regions by adaptively adjusting weights across spatial locations and channels, enabling the model to capture both critical features and contextual information.

In the field of image steganalysis, attention mechanisms have also been explored. For instance, Zhang et al. [23] proposed a multi-attention strategy combining spatial and channel attention to improve detection accuracy and better exploit the texture information of the image. Ma et al. [24] introduced LGS-Net, which used synergetic attention in steganalysis networks, focusing on both global structures and local details to enhance the model's ability to capture steganographic signals. Hu et al. [25] proposed a novel self-adaptive steganalysis method based on visual attention and deep reinforcement learning. By selecting regions of interest through visual attention and performing successive decisions via reinforcement learning, the model generates compact summary regions to improve the training set quality and overall detection performance. Bravo-Ortiz a et al. [26] proposed a Convolutional Vision Transformer (CVT) for spatial domain image steganalysis. This method combines the local perceptual ability of convolutional operations with the global modeling advantage of attention mechanisms, and is able to capture both local details and global dependencies in an image, thus significantly improving the classification accuracy of steganalysis. Guo et al. [27] designed a steganographic feature extraction algorithm TSNet based on dual-stream networks and introduced a fusion module, DCTAM-FM, which incorporates the Discrete Cosine Transform Attention Mechanism (DCTAM). By utilizing DCTAM for fusion and feature enhancement of two branches in a dual-stream network, the accuracy of steganography detection was significantly improved.

These developments demonstrate that deep learning significantly improves the detection accuracy of image steganalysis compared to traditional methods. However, despite its advantages, several challenges remain in practical applications. Deep models often struggle to effectively focus on subtle steganographic perturbations, key features are easily diluted through network depth, and critical information may be lost during training, ultimately impairing detection accuracy. Moreover, current research provides limited exploration into strategies for focusing on steganography-sensitive regions and mitigating feature loss. Detection accuracy on public benchmark datasets still leaves room for improvement, and the overall capacity for steganographic feature extraction remains constrained by apparent bottlenecks.

3 Methods

To address the challenges of accurately focusing on steganography-sensitive regions and minimizing the loss of steganographic information, a novel network architecture is proposed in this paper. Its detailed structure, as shown in Fig. 1, the proposed model consists of three main components: a preprocessing stage, a feature extraction stage, and a classification stage. Each component is elaborated in the following subsections.

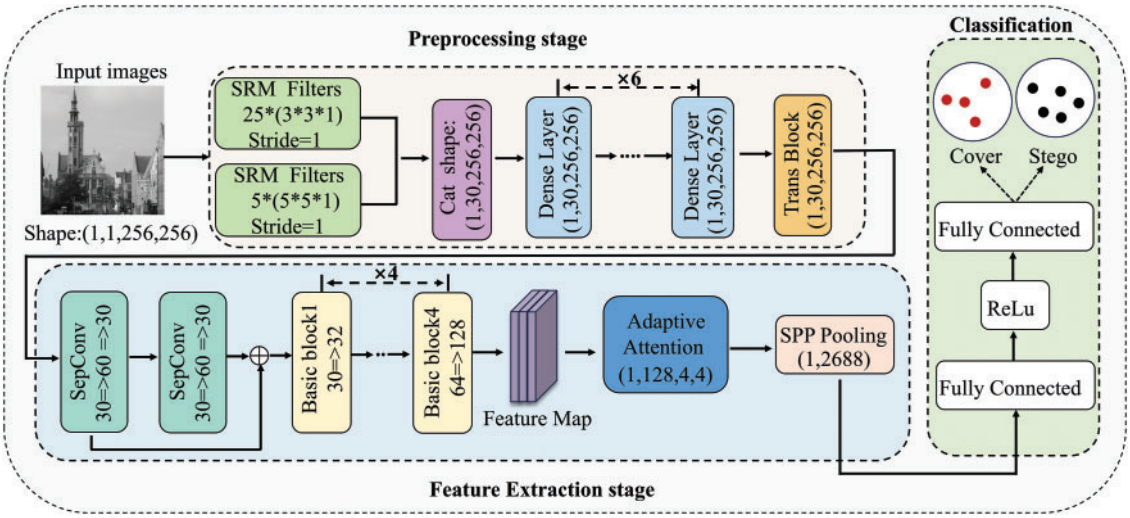

Figure 1: Proposed network architecture

The input image first undergoes a preprocessing stage to extract features. Next, it proceeds to the feature extraction stage for feature transformation. Finally, the transformed feature map is passed to the classification stage to obtain the result.

As shown in Table 1, the specific dimensional changes of the input image during the modeling process are provided in this paper. DenseLite is a Lightweight DenseNet module consisting of 6 dense layers and 1 Transblock.

Table 1: Changes in the input feature maps within the model

	Specific module	Input size	Output size
Preprocessing	SRM	(1, 1, 256, 256)	(1, 30, 256, 256)
	DenseLite	(1, 30, 256, 256)	(1, 30, 256, 256)
Feature extraction	SepConv	(1, 30, 256, 256)	(1, 30, 256, 256)
	Basick block 1	(1, 30, 256, 256)	(1, 32, 64, 64)
	Basick block 2	(1, 32, 64, 64)	(1, 32, 16, 16)
	Basick block 3	(1, 32, 16, 16)	(1, 64, 4, 4)
	Basick block 4	(1, 64, 4, 4)	(1, 128, 4, 4)
	Adaptive attention	(1, 128, 4, 4)	(1, 128, 4, 4)
	SPP	(1, 128, 4, 4)	(1, 2688)
Classification	Linear	2688	1024
	ReLu	1024	1024
	Linear	1024	2

3.1 Preprocessing Stage

SRM: In the field of steganalysis, steganographic images are visually nearly identical to their corresponding cover images, and the embedded steganographic signals are typically very weak. These weak signals are easily overshadowed by the natural image content, which interferes with effective feature extraction. To address this challenge, researchers have introduced the Spatial Rich Model (SRM) technique. The core idea of

SRM is to transform the image from the spatial domain to the residual domain, where steganographic noise can be more effectively emphasized, thereby creating more favorable conditions for subsequent analysis and detection. However, while traditional fixed SRM filters have proven effective in hand-crafted feature-based approaches, they present limitations when applied within deep learning frameworks. Literature [28] points out that traditional fixed SRM filters cannot be fully optimized and adapted in deep networks, which largely limits their feature extraction capabilities and makes it difficult to fully utilize the potential of deep learning models. In view of this, this paper improves the traditional SRM.

Specifically, as shown in Fig. 1, Unlike ZhuNet, ZhuNet initializes 30 SRM-inspired high-pass filters (including 25 hand-crafted 3 × 3 SRM filters and 5 manually-designed 5 × 5 filters) at the preprocessing layer, and then trains them in the optimization process. Specifically, the SRM, proposed in this paper, change them into 25 convolutional kernels of size 3 × 3 and 5 kernels of size 5 × 5, all of which are optimized by end-to-end training. This state-of-the-art design allows the module to efficiently capture subtle hidden noise patterns through adaptive learning, rather than relying on expertly crafted filters. In addition, the use of multi-scale filters enables the extraction of high-frequency signals at various spatial resolutions, which is crucial for detecting various em-bedding strategies. Instead of summing the filter outputs or selecting a fixed response, the proposed module employs channel concatenation to fuse all feature maps, resulting in a richer and complementary representation while preserving subtle information. Compared to ZhuNet's fixed SRM setup, this dynamic, multi-scale approach not only enhances the flexibility and expressiveness of the pre-processing stage, but also lays a more robust feature foundation for downstream steganography analysis. Each kernel produces a feature map, and all 30 feature maps are concatenated along the channel dimension. This process can be represented by Eq. (1).

$$Output = Concat\left(f_{3\times3}^{25}(x), f_{5\times5}^{5}(x)\right) \tag{1}$$

DenseLite: Since the steganographic features processed by SRM in the preprocessing stage are critical, the goal is to retain as much feature information as possible. In image steganalysis, the strength of DenseNet [16] lies in its dense connectivity, which effectively aggregates both low-level and high-level features, making it particularly valuable for fine-grained steganographic feature extraction. Although DenseNet performs exceptionally well, its complex network structure, large number of parameters, and high computational cost can be a limitation. In this paper, integrate the core principles of DenseNet and design a simplified module that preserves its key advantages while reducing computational cost and model complexity, making it more suitable for image steganalysis tasks. This approach enables us to enhance the extraction of steganographic features while maintaining the model's efficiency and scalability.

As shown on the left side of Fig. 2, conventional residual networks (e.g., ResNet) propagate information by summing the output of a previous layer with that of the current layer. While this additive shortcut mechanism effectively mitigates gradient vanishing, it constrains the network's ability to retain diverse and fine-grained features, particularly in deeper architectures, where accumulated information may become increasingly diluted. In contrast, DenseLite adopts a more effective strategy by densely concatenating the outputs of all preceding layers along the channel dimension. This dense connectivity allows each layer to directly access a richer set of contextual representations, thereby preserving fine-grained features and reducing information loss in the early stages of the network. Compared with traditional methods such as ZhuNet and SRNet, DenseLite enhances feature reuse and facilitates more efficient information flow. These advantages not only improve the model's capacity to retain detailed information but also ensure better preservation of the input signal throughout the network, resulting in superior performance in complex steganalysis tasks. As shown in the right half of Fig. 2, DenseLite consists of two main components: layers and TransBlock. The specific process is as follows:

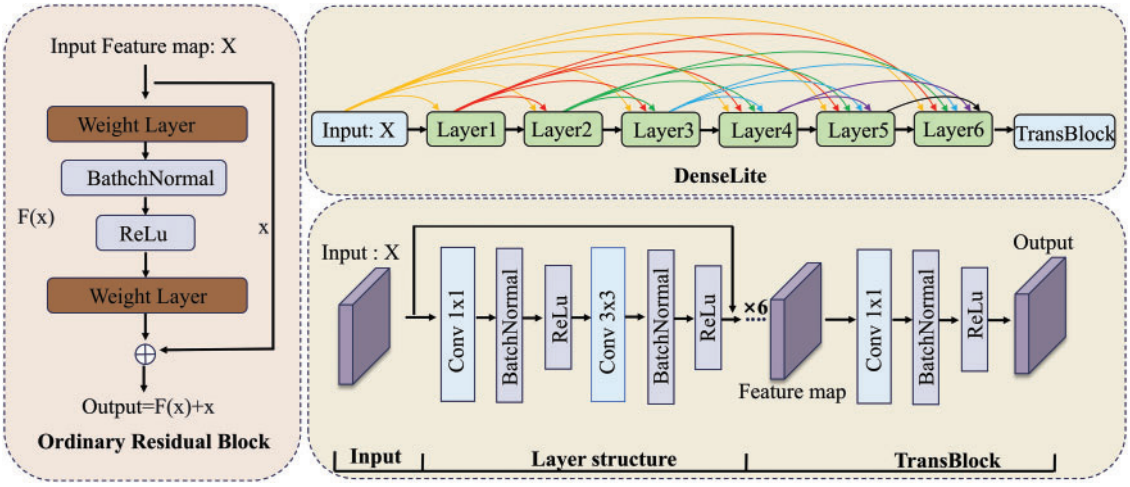

Figure 2: The ordinary residual network and proposed DenseLite network architecture. Denselite's layers, linked by red, yellow, green, blue, and purple lines, each containing all of the previous information. The layer structure represents the internal structure of a layer. Transblock represents a transition layer

The SRM-processed feature map is input. $X_0 \in R^{B\times C_0\times H\times W}$ Into dense layers, and the input of each dense layer consists of all the outputs of the previous layer. The input of the lth layer is $[X_0, X_1, \ldots\ldots, X_{l-1}]$, the spliced features are nonlinearly transformed $H_l(\cdot)$ to generate a new feature X_l For the current layer. H_l Includes Batchnormal and ReLu, as shown in Eq. (2).

$$X_l = H_l\left(Concat[X_0, X_1, \ldots\ldots, X_{l-1}]\right) \tag{2}$$

Then the feature maps that have been changed in the six Dense layers are input to the next layer through the TransBlock connection, as shown in Eq. (3).

$$\text{TransBlock} = H_l\left(Conv_{1\times1}(x)\right) \tag{3}$$

Furthermore, as shown in Table 2, due to DenseLite's feature reuse and parameter sharing, the number of parameters in model framework does not increase significantly.

Table 2: Number of network parameters for mainstream methods

Method	SR-Net	Zhu-Net	Proposed
Number of parameters	4.78 million	2.87 million	3.23 million

3.2 Feature Extraction Stage

To further improve image steganography detection capabilities, the main components in the feature extraction stage are: depthwise separable convolution, four base convolutions, Adaptive attention mechanism, and SPP pyramid pooling. Each structure is explained separately below.

Depthwise Separable Convolution: to facilitate further extraction of features, depthwise separable convolution is used [29], which is named "SepConv" and can be categorized into Deep Convolution and Point-by-Point Convolution. The combination of the two enhances the network's ability to learn hidden

written information. As shown in the upper part of Fig. 3, ordinary convolution is a convolution operation performed on all channels of the input data at the same time, and the convolution kernel slides over both the spatial dimensions (height and width) and the channel dimensions, mixing the information of all channels. As shown in the lower part of Fig. 3, Depthwise separable convolution, each channel of the input feature map is individually convolved by a 3 × 3 convolution kernel and then spliced. The spliced feature maps are then subjected to a 1 × 1 convolution operation kernel and finally output, as shown in Eq. (4).

$$Conv_{sep}(F) = Pointwise(Depthwise(F)) \tag{4}$$

Although the depth-separated convolution can reduce the number of parameterizers, it also brings some information loss. Therefore, the input and output are fused by residual linking [30], as shown in Eq. (5).

$$\text{Output} = F + Conv_{sep2}\left(Conv_{sep1}(F)\right) \tag{5}$$

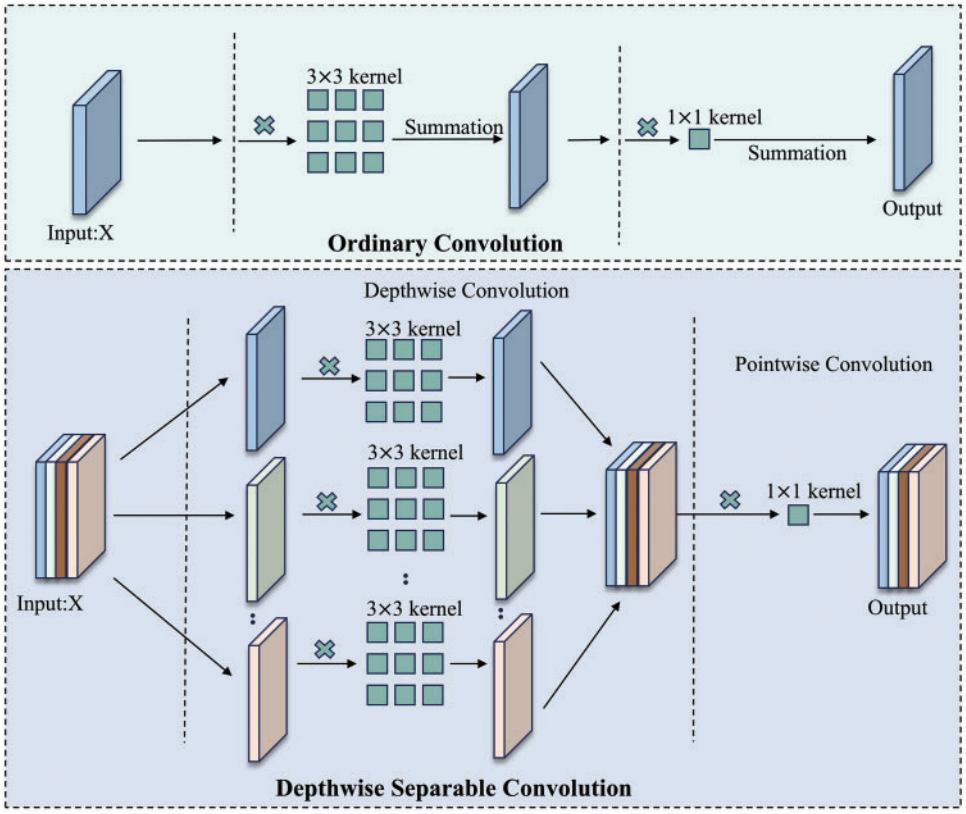

Figure 3: The process of ordinary convolution and depthwise separable convolution

Adaptive Attention Mechanism: To enable the model to focus on key steganographic regions, inspired by the approach of Woo et al. [22,31], propose an improved dual-branch attention design named adaptive attention, as illustrated in the right part of Fig. 4. Unlike conventional CBAM-like modules that sequentially apply channel and spatial attention with fixed weighting, the method introduces a gated fusion strategy that dynamically learns the contribution of each attention branch. Specifically, channel and spatial attention are computed in parallel and then jointly fused through a learnable gating module, which adaptively balances

their influence based on the input content. This design brings two key innovations compared to traditional dual-attention mechanisms: mutual guidance, where the channel and spatial attention modules share and respond to the same intermediate features, promoting cross-branch interaction and enhancing focus on steganography-sensitive regions, and dynamic fusion, where a learned gate modulates the final attention map, allowing the model to adaptively emphasize the more informative branch under varying conditions, thus improving robustness across different payloads and embedding patterns. The process can be expressed as follows:

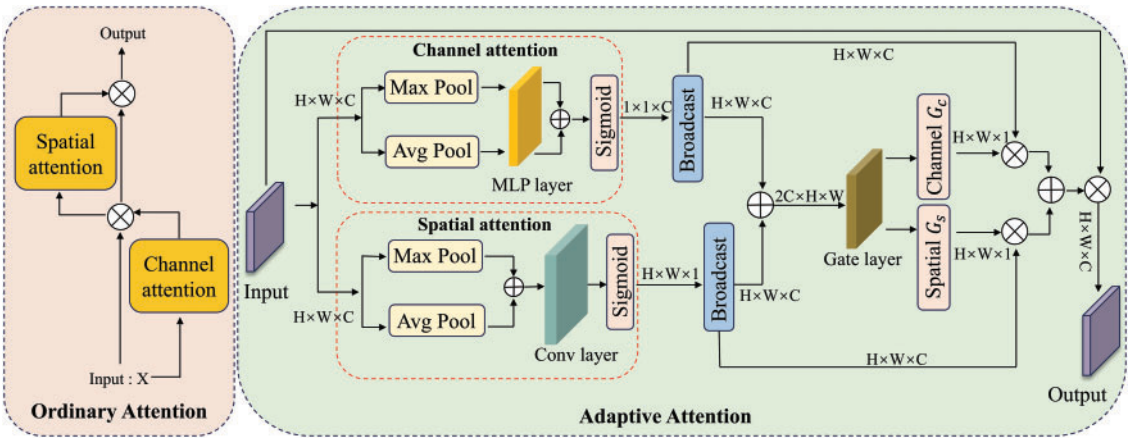

Figure 4: Ordinary hybrid attentional mechanism and proposed adaptive attention mechanism

(1) Compression Part: To preserve or suppress features at different spatial and channel locations, the feature maps are first input into the channel attention module and the spatial attention module, respectively. The channel and spatial attention mechanisms generate their respective feature maps using maximum pooling and average pooling. The channel attention map is then obtained by passing through the MLP layer, followed by concatenation and σ (Normalizing the importance of each channel to [0, 1] allows the model to selectively suppress or enhance certain channel features). The spatial attention features are concatenated along the channel dimension, and the spatial attention map is extracted using a Conv layer ($f^{7\times7}$) and σ (Compresses the significance of each spatial location of the output to [0, 1]). The final result is expressed as follows: the channel attention features are $M_c \in R^{B\times C\times1\times1}$, the spatial attention features are $M_s \in R^{B\times1\times H\times W}$, as shown in Eqs. (6) and (7).

$$M_c = \sigma\left(\left(MLP\left(AvgPool\left(F\right)\right)\right) + MLP\left(MaxPool\left(F\right)\right)\right) \in R^{B\times C\times1\times1} \tag{6}$$

$$M_s = \sigma\left(f^{7\times7}\left(\left[AvgPool_c\left(F\right) + MaxPool_c\left(F\right)\right]\right)\right) \in R^{B\times1\times H\times W} \tag{7}$$

Then $M_c \in R^{B\times C\times1\times1}$, and $M_s \in R^{B\times1\times H\times W}$ are broadcasted and expanded to the same shape $B \times C \times H \times W$ for element-wise multiplication with the input feature map, which are denoted as $M_c^{'}$ and $M_s^{'}$, respectively, as shown in Eqs. (8) and (9).

$$M_c^{'} = M_c \times 1_{H\times W} \in R^{B\times C\times H\times W} \tag{8}$$

$$M_s^{'} = M_s \times 1_C \in R^{B\times C\times H\times W} \tag{9}$$

(2) Interaction Part: To further extract global attention features and enable the network to deploy the attention region more flexibly. Firstly, $M_c^{'}, M_s^{'}$ are spliced into M_{concat}, as shown in Eq. (10).

$$M_{concat} = Concat\left(M_c^{'}, M_s^{'}\right) \in R^{B\times 2C\times H\times W} \tag{10}$$

The features are then fed into the Gate layer, which automatically learns the preference weights of each spatial location for the channel and spatial attention through the end-to-end backpropagation process. W_1 represents the weight of the first convolutional layer, responsible for extracting the fused intermediate feature representation. W_2 represents the weight of the second convolutional layer, which compresses the intermediate representation into two channels. The weights are then mapped to the (0, 1) range using σ(Ensure that the generated gating coefficients fall in [0, 1] to rationally weight channel attention and spatial attention) to enable weighted control. These two channels represent the preference weights of channel attention and spatial attention at each spatial location, respectively, as shown in Eq. (11).

$$G = G\left(M_{concat}\right) = \sigma\left(W_2 * ReLu\left(W_1 * M_{concat}\right)\right) \in R^{B\times 2\times H\times W} \tag{11}$$

After gating adaptive learning, the final segmentation outputs two gating weights G_c, G_s, as shown in Eq. (12).

$$\left[G_c, G_s\right] = \text{Split}\left(G, dim = 1\right) \tag{12}$$

(3) Fusion Part: To modulate the different branches of attention, the effective parts are amplified, and the ineffective parts are suppressed. The obtained G_c and G_s are applied to the channel's attention $M_c^{'}$ and spatial attention $M_s^{'}$. To obtain new spatial and channel feature maps, respectively, which are then summed. Finally, the newly fused attention maps are multiplied with the original input to obtain the final output features, as shown in Eqs. (13) and (14).

$$M_{fused} = G_c \cdot M_c^{'} + G_s \cdot M_s^{'} \tag{13}$$

$$F_{output} = F_{input} \cdot M_{fused} \tag{14}$$

SPP Pyramid Pooling: To better handle the feature maps from the fused adaptive attention mechanism, as shown in Fig. 5, SPP pyramid pooling is proposed to be used [18]. It helps to improve the accuracy of the model for the steganalysis task. In this paper, set the window sizes of 4 × 4, 2 × 2, and 1 × 1, and perform multi-scale pooling in parallel to fuse local features with global features of different granularities to form a joint representation of multi-level sensory fields, and the final spatial pyramid pooling output is (1 + 4 + 16) × 128 dimensions. The specific structure of spatial pyramid pooling can be represented as shown in Eq. (15).

$$y_{spp} = Concat\left(Pool_{1\times 1}\left(x\right), Pool_{2\times 2}\left(x\right), Pool_{4\times 4}\left(x\right)\right) \tag{15}$$

x is the input feature map, $Pool_{k\times k}$ denotes the pooling operation after dividing the feature map into k × k subregions, and $Concat$ denotes the splicing of pooling results of different scales.

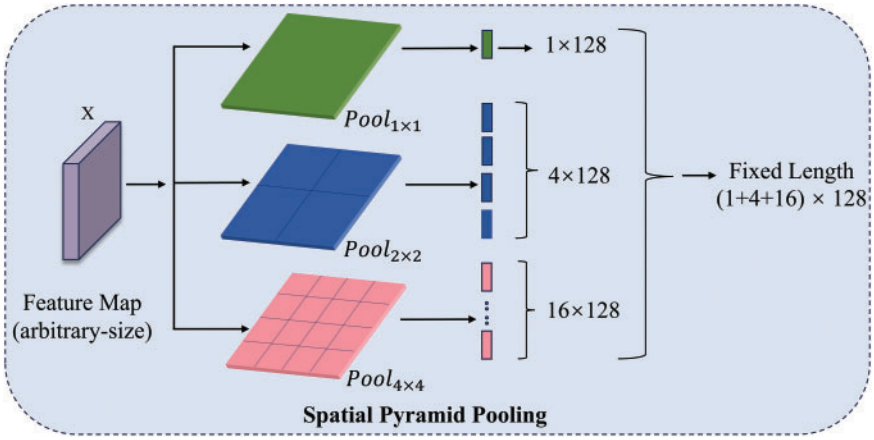

Figure 5: Detailed structure of spatial pyramid pooling

3.3 Classification Stage

Feature vectors pass through the SPP, which are then used as inputs for the final classification. The classification module, shown in Fig. 1 consists of two fully connected layers and an activation function, with the final predicted classification being labeled as 'cover' or 'stego'. The specific process for the input and output units is as follows: the first fully connected layer maps the input dimension from (2688) to (1024), followed by a ReLu activation function to enhance the nonlinear representation. The second fully connected layer further maps the features from (1024) to (2), representing the final classification output.

4 Experiments

In this section, the relevant experimental parameters are introduced. Comparisons are also made with state-of-the-art image steganalysis methods, as well as a large number of ablation experiments are set up to verify the veracity and effectiveness of method.

4.1 Datasets and Implementation Details

Two benchmark datasets: Bossbase 1.01 and BOWS2 are used to evaluate the detection performance of network architectures. In this paper, refer to BOSSBase 1.01 as BOSSBase for short, where BOSSBase contains 10,000 grayscale images of size 512 × 512 from seven different cameras, and BOWS2 contains 10,000 grayscale images of size 512 × 512, which cover a variety of scenes and objects, such as street scenes, indoor environments, natural landscapes, and objects. To reduce the difficulty of the experiment, use the imresize function in MATLAB to convert both to 256 × 256. Four classical image steganography algorithms, including WOW, S-UNIWARD, and HILL, are used in the experiment to evaluate the performance of the proposed network architecture. For each steganography algorithm, each embedding rate can produce 10,000 pairs of cover-target images, and four embedding rates of 0.1, 0.2, 0.3, and 0.4 bpp are chosen in this paper. The dataset is divided into training, testing, and validation sets in the ratio of 4:5:1 with no duplicate data in the four samples. Data augmentation of the training data helps to improve the generalization ability of the model, enabling better adaptation to different environments. In this paper, 10,000 grayscale images from the BOWS2 dataset are incorporated into the existing training set. Eventually, the training set consists of 10,000 grayscale images from the BOWS2 dataset and 4000 grayscale images from the BOSSbase dataset, totaling 14,000 images. On the other hand, the validation set consists of 1000 grayscale images from the BOSSbase dataset, while the test set consists of 5000 grayscale images from the same dataset.

To train the network, a total of 200 epochs were performed with a batch size of 16. The AdamW optimizer was used, with β1 = 0.9 and β2 = 0.999. The initial learning rate was 0.001 and the weight decay was 0.0005. At the 80th, 140th and 180th epochs, the learning rate was divided by 10. To ensure fusion stability and to prevent overfitting at the later stages of the training, a data To ensure fusion stability and prevent overfitting at later stages of training, data enhancement techniques were used, including randomized horizontal flips and 90-degree rotations. In addition, batch normalization (BN) was performed using PyTorch's default settings. To ensure consistency and fairness of comparisons, all experiments were conducted on an NVIDIA RTX 3090 GPU with 24 GB of RAM, using PyTorch version 1.11.0 and Python 3.8.

4.2 Comparative Study with Related Methods

The effectiveness of the proposed method will be validated by comparing it with other related steganalysis methods. Extensive experiments will be conducted on publicly available datasets, evaluating performance in terms of detection accuracy. The experimental results will demonstrate the superiority of method over existing approaches and highlight its potential for practical steganalysis applications. As shown in Fig. 6, the figure visualizes the performance of the proposed model under the training and validation accuracy curves for WOW 0.4 bpp and S-UNIWARD 0.4 bpp conditions, which further proves its effective extraction capability and good adaptability to steganographic features in image steganalysis tasks.

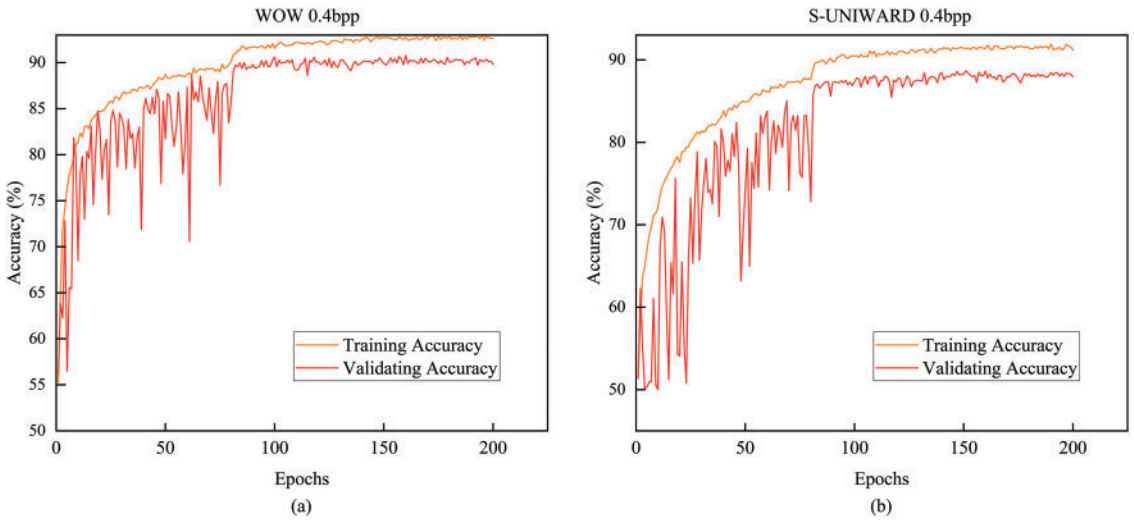

Figure 6: During the training process at 0.4 bpp: **(a)** the changes in the training set and validation set of WOW; **(b)** the changes in the training set and validation set of S-UNIWARD

As shown in Fig. 7, the curves for the methods proposed in this paper rise faster than those of the other models, indicating a more rapid improvement in detection accuracy over time. This suggests that the method not only achieves high performance but also does so more efficiently compared to other models. The rapid rise in accuracy shown in the curves and the consistent improvements across different algorithms and embedding rates demonstrate that the model is not only more effective but also more adaptable to various steganographic techniques. This is further validated by the sharp contrast with other methods, which shows that the approach offers superior performance both in terms of accuracy and robustness. The results underline the advantages of the adaptive attention mechanism and DenseLite architecture in handling diverse steganography challenges. The methods in this paper show an overall silly lead compared to the best methods of the day.

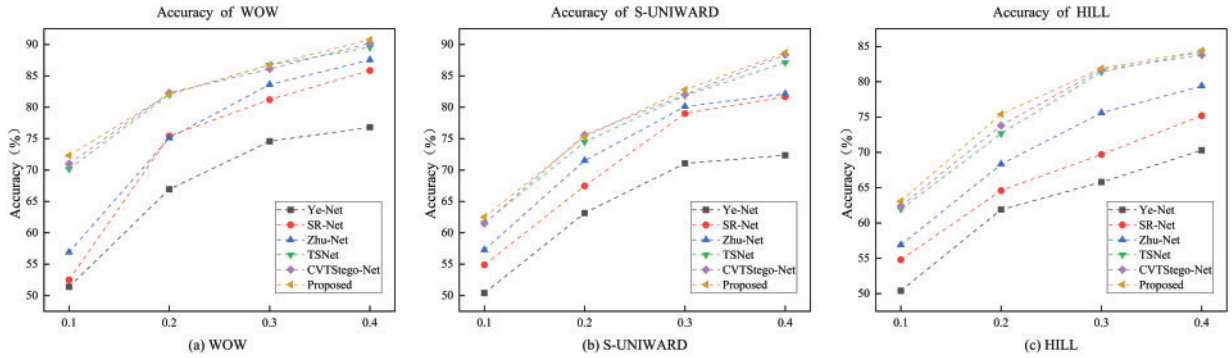

Figure 7: Comparison of the accuracy of various methods under BOSSBase. **(a–c)** Plots of the change in accuracy of WOW, S-UNIWARD, and HILL at four bpp, respectively

From Table 3, for instance, with the 0.4 bpp WOW algorithm, the accuracy of the proposed method reaches 90.77%, which is 13.95%, 4.91%, 3.24%, 1.21%, and 0.56% higher than Ye-Net, SR-Net, Zhu-Net, TSNet, and CVTStego-Net, respectively. Similarly, using the S-UNIWARD algorithm (0.4 bpp), the method outperforms the other models by 16.34%, 7.01%, 6.56%, 1.57%, and 0.33%, respectively. Using the HILL algorithm at 0.4 bpp outperforms other methods by 14.10%, 9.20%, 5.00%, 1.15%, and 0.60%, respectively. The method outperforms other models on other embedding rate data, and the proposed method also significantly outperforms other models. This shows that the deep steganalysis method based on the adaptive attention mechanism and DenseLite can perform the steganalysis task well. This not only proves that the model performs well in terms of accuracy, but also shows that it is significantly robust under different steganography algorithms and embedding rates, which proves its generalizability in real-world application scenarios.

Table 3: Comparison of the accuracy (%) of the method with Ye-Net, SR-Net, Zhu-Net, TSNet, and CVTStego-Net under BOSSBase

Steganography	Payload (bpp)	Method					
		Ye-Net	SR-Net	Zhu-Net	TSNet	CVTStego-Net	Proposed
WOW	0.1 bpp	51.40	52.50	56.90	70.17	71.00	72.32
	0.2 bpp	66.94	75.42	75.10	82.05	82.30	82.09
	0.3 bpp	74.58	81.22	83.62	86.10	86.40	86.76
	0.4 bpp	76.82	85.86	87.53	89.56	90.21	90.77
S-UNIWARD	0.1 bpp	50.40	54.90	57.30	61.38	61.49	62.51
	0.2 bpp	63.15	67.47	71.51	74.47	74.60	75.29
	0.3 bpp	71.06	79.01	80.12	81.86	82.02	82.78
	0.4 bpp	72.35	81.68	82.13	87.12	88.36	88.69
HILL	0.1 bpp	50.40	54.80	56.90	61.97	62.40	63.10
	0.2 bpp	61.90	64.60	68.33	72.67	73.80	75.41
	0.3 bpp	65.80	69.70	75.60	81.41	81.70	81.90
	0.4 bpp	70.30	75.20	79.40	83.25	83.80	84.40

As shown in Table 4, to enhance the robustness and generalization capability of the model, employed data augmentation strategies, including random 90-degree rotations and horizontal flipping with a 50%

probability during training. These techniques were applied to the expanded dataset of Bossbase and Bosw2 for comparison. The augmentations not only increased the diversity of the data but also improved the model's ability to focus on subtle steganographic patterns that are invariant to spatial transformations. Under these conditions, the proposed method achieved remarkable results, especially at the low embedding rate of 0.2 bpp, which is typically considered the most challenging due to minimal embedding distortion. Under the WOW algorithm at 0.2 bpp, the method achieved the highest accuracy of 87.09%, outperforming Ye-Net by 15.05%, SR-Net by 3.67%, Zhu-Net by 3.53%, TSNet by 1.04%, and even CVTStego-Net by 0.51%. This significant improvement demonstrates the effectiveness of model design in extracting weak and imperceptible steganographic signals that are often overlooked by conventional convolutional architectures. Similarly, with the S-UNIWARD algorithm at 0.2 bpp, the method obtains an accuracy of 81.79%, which is 11.94% higher than Ye-Net, 3.06% higher than SR-Net, 2.17% higher than Zhu-Net, 0.32% higher than TSNet, and 1.09% higher than CVTStego-Net. These results show that the proposed model not only generalizes well to different embedding algorithms but also remains efficient even under subtle embedding conditions. These results indicate that the proposed model not only generalizes well across different embedding algorithms but also remains highly effective even under subtle embedding conditions. This versatility demonstrates the robustness of the model in various scenarios. It suggests that the model can be reliably applied to diverse datasets and tasks without significant performance degradation. Furthermore, the ability to maintain effectiveness under subtle embedding conditions highlights its potential for real-world applications where data nuances are common.

Table 4: Comparison of the accuracy (%) of the method with Ye-Net, SR-Net, Zhu-Net, TSNet, and CVTStego-Net under BOSSBase and BOWS2

Method	WOW		S-UNIWARD	
	0.2 bpp	0.4 bpp	0.2 bpp	0.4 bpp
Ye-Net	72.04	80.27	69.35	78.69
SR-Net	83.42	91.03	78.23	89.61
Zhu-Net	83.56	91.87	79.62	90.50
TSNet	86.05	92.66	81.47	91.12
CVTStego-Net	86.58	93.80	80.70	90.45
Proposed	87.09	93.54	81.79	90.59

4.3 Ablation Studies

To verify the effect of the SRM preprocessing method on the detection performance of steganalysis, a combination of fixed SRM and customized SRM filters is used as the preprocessing module. As shown in Table 5, the proposed method achieves the highest detection accuracies of 90.77% and 88.69% at a 0.4 bpp load for WOW and S-UNIWARD, respectively. Compared to the fixed SRM scheme, the accuracy is improved by 1.69% and 1.49%, respectively. This result suggests that introducing a combination of filters with different sizes and structures can more effectively extract steganographic features from the image and enhance the model's ability to detect weak signals, compared to using a fixed set of 30 identical filters.

Table 5: Comparison of the accuracy (%) of the Fixed SRM and the SRM proposed in this paper

Method	Fixed SRM	Proposed
WOW 0.4 bpp	89.08	90.77

(Continued)

Table 5 (continued)

Method	Fixed SRM	Proposed
S-UNIWARD 0.4 bpp	87.20	88.69

To investigate how information loss can be effectively reduced during the early stages of steganalysis, this study introduces the DenseLite module into the preprocessing phase. Table 6 presents a comparative evaluation of model performance with and without DenseLite convolution. The results indicate a substantial drop in detection accuracy—5.5% for WOW and 7.1% for S-UNIWARD—when DenseLite is excluded. This demonstrates the module's ability to better preserve steganographic features extracted by high-pass filtering. Compared to conventional convolutional layers commonly used in previous works, such as YeNet and SRNet, which typically apply standard residual or plain convolution blocks, DenseLite leverages dense connectivity to enhance low-level feature reuse and promote uninterrupted information flow across layers. This dense design not only alleviates the degradation of subtle steganographic signals but also allows the model to retain more semantically relevant details during feature transformation. Furthermore, the lightweight nature of DenseLite minimizes unnecessary parameter growth, avoiding overfitting while still preserving discriminative patterns. These advantages collectively contribute to a stronger representation capacity and improved robustness in steganalysis tasks.

Table 6: Validating the accuracy (%) of DenseLite under WOW and S-UNIWARD

Method	No DenseLite	DenseLite
WOW 0.4 bpp	85.27	90.77
S-UNIWARD 0.4 bpp	81.59	88.69

Experiments were conducted with a gradual increase in the number of layers from 1 to 7 to explore the optimal layer configuration. As shown in Fig. 8, the model performs optimally with an accuracy of 90.77% when the number of DenseLite layers is six. However, when the number of layers is increased to seven, the model performance decreases, likely due to the introduction of redundant information and the increased computational and storage costs, which reduce training and deployment efficiency.

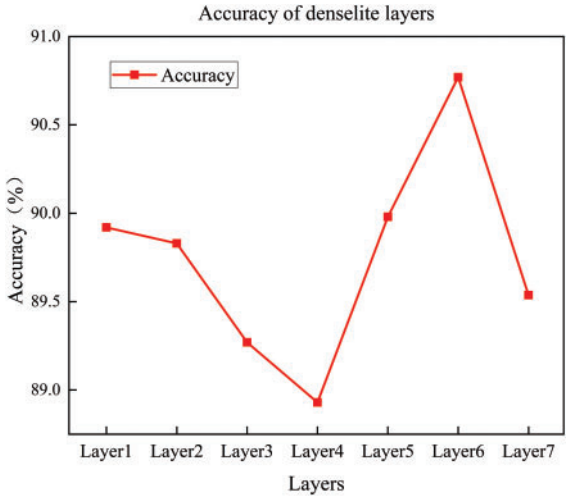

Figure 8: Accuracy (%) for different layers of DenseLite

To evaluate the effectiveness of the proposed adaptive attention mechanism, conducted ablation experiments under different attention module configurations and fusion strategies. Table 7 lists the performance comparison under different embedding rates. The proposed adaptive attention mechanism consistently maintains the best performance with a maximum accuracy of 90.77%. At higher embedding rates (0.4 bpp), the difference in detection accuracy between adaptive attention and no-attention baseline is relatively small (0.79% and 1.54%, respectively).

Table 7: Accuracy (%) under different combinations of attention

Method	WOW		S-UNIWARD	
	0.2 bpp	0.4 bpp	0.2 bpp	0.4 bpp
NO attention	78.64	89.98	72.56	87.15
Spatial attention	80.15	88.79	73.88	87.05
Channel attention	76.24	85.85	65.32	82.49
CBAM attention	78.35	87.59	72.19	86.46
Proposed	82.09	90.77	75.29	88.69

However, when the embedding rate was lower (0.2 bpp), the cryptographic signals were even weaker, and the difference widened significantly to 3.45% and 2.73%, indicating that adaptive attention was more capable of capturing weak signals.

The performance is significantly improved when using only spatial attention, especially at low embedding rates, where the accuracy gap narrows to 1.94% and 1.41%, respectively. In contrast, the lowest performance is achieved when using only channel attention, with accuracy gaps of 4.92% and 6.20% at 0.4 bpp and 5.85% and 9.97% at 0.2 bpp, respectively. In addition, use the CBAM-style channel spatial attention. While this approach outperforms single-branch attention, the accuracy gap is still above 2%–3%, suggesting that it fails to fully utilize the complementary nature of the two branches due to its static fusion nature.

In contrast, the proposed adaptive attention mechanism introduces a gated fusion strategy that dynamically adjusts the contributions of channel and spatial attention based on the input feature characteristics. This learned attention fusion enables more flexible and effective focus on subtle steganographic features, especially under low payload conditions, thereby significantly enhancing the model's discriminative ability.

In this experiment, four different attentional adaptive methods are compared to validate the effectiveness of the proposed. G_{net} method. Namely, Fixed α (fixed fusion weights), Dynamic (dynamic convolution), MLP, and G_{net}.

From Table 8, it can be concluded that the G_{net} strategy proposed in this paper has the most superior performance with WOW 0.4 bpp and S-UNIWARD 0.4 bpp reaching 90.77% and 88.69%, respectively. However, the fixed α, which employs a fixed proportion of weighting for channel attention and spatial attention, is 1.98% and 1.59% lower than G_{net}, respectively; the dynamic method adjusts the fusion weights by dynamic convolution to extract the statistical information of the input image to adjust the fusion weights, which are 1.49% and 1.89% lower than G_{net}, respectively; and the MLP method, which introduces nonlinear modeling to predict the weights from the shallow perceptual structure, is 0.65% and 1.37% lower than G_{net}, respectively.

Table 8: Comparison of the accuracy (%) of the adaptive weighting method proposed in this paper with others

Method	Fixed α	Dynamic	MLP	G_{net}
WOW 0.4 bpp	88.79	89.28	90.12	90.77
S-UNIWARD 0.4 bpp	87.10	86.80	87.32	88.69

Therefore, G_{net} uses a lightweight gated network, which effectively enhances the model's ability to model the importance of different features by introducing a gating mechanism to dynamically adjust the fusion ratio of channel and spatial attention. Under different loads and different embedding algorithms, the G_{net} The fusion strategy achieves the best performance.

To validate the effect of multiple weight assignments before and after fusion. Before reshaping the shapes, the effects of assigning two weights to spatial and channel attention are compared. NO σ is to use sigmoid only once after fusion. Spatial σ is to use sigmoid only for spatial branching, channel σ is to use sigmoid only for channel branching, and the proposed is to use sigmoid for both space and channel.

As shown in Fig. 9, the proposed method consistently outperforms the other approaches, exhibiting regionally flat and stable loss curves as well as training accuracies. This indicates that the model achieves a steady and reliable learning process. In contrast, the other three methods display significant fluctuations in both loss values and accuracy throughout the training process, with particularly notable instability at the WOW 0.4 bpp setting. The losses for these methods are erratic, which suggests that they struggle to converge and may be prone to overfitting or underfitting during training. On the other hand, the method proposed in this paper demonstrates a gradual stabilization of the training process. This improvement in stability is indicative of the model's robustness, allowing it to learn more efficiently and avoid unnecessary oscillations in performance. Over time, the proposed method shows a steady increase in accuracy and a loss reduction, highlighting the effectiveness of the approach in achieving both high performance and stable learning dynamics across different steganographic settings. This stability is essential for real-world applications, where consistency and reliability are crucial factors in the success of steganalysis systems.

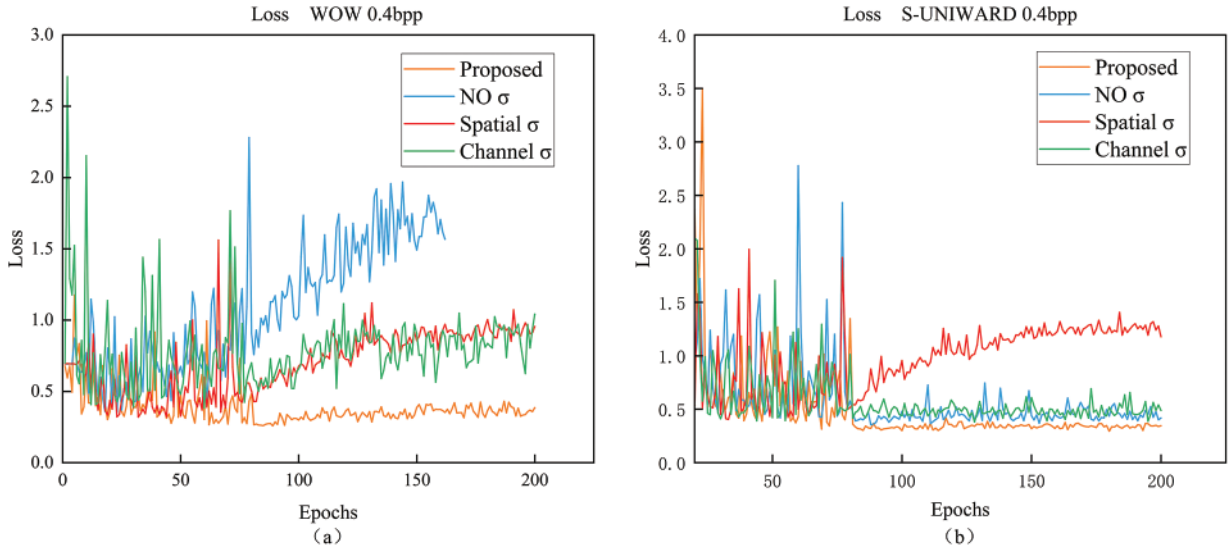

Figure 9: (Continued)

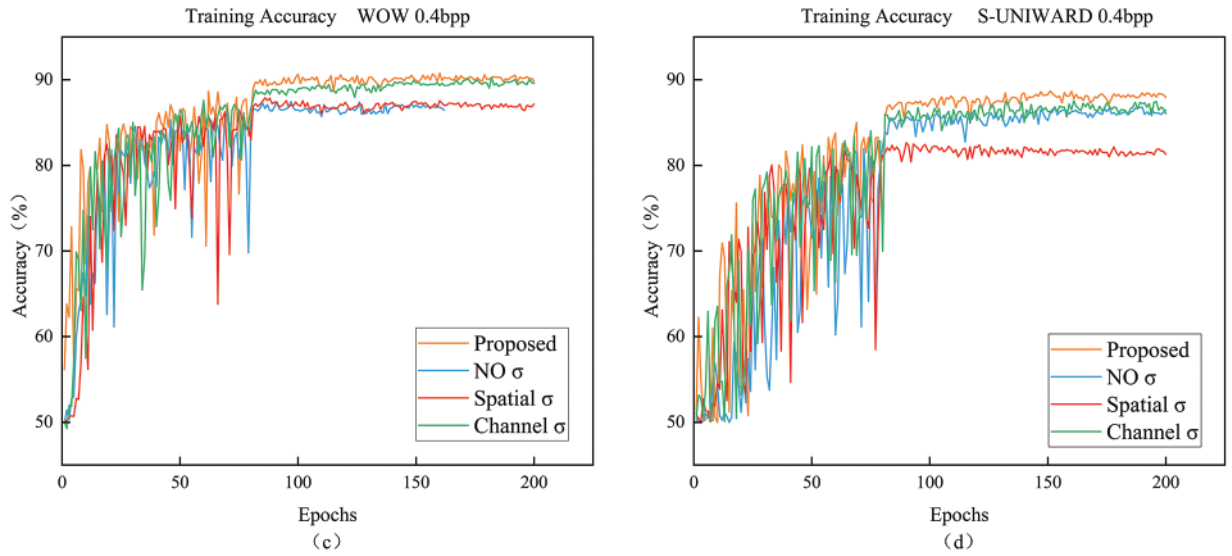

Figure 9: Whether or not to apply the sigmoid function multiple times: (**a,b**) show the change in loss during training on the validation set; (**c,d**) show the change in accuracy (%) during training on the validation set

As shown in Table 9, the methods proposed in this paper are both optimal up to 90.77% and 88.69%. The NO σ is lower by 3.32% and 1.84% compared to the methods proposed in this paper. The effect of spatial σ is only 2.88% lower under WOW, but drops significantly to 6.06% under S-UNIWARD, which suggests that spatial attention suffers from insufficient generalization in some steganography algorithms. In contrast, channel σ brings a significant performance improvement with a difference of only 0.5% and 1.15% in WOW and S-UNIWARD, which suggests that channel attention is more helpful in capturing semantic information related to steganographic functions.

Table 9: Accuracy comparison under WOW and S-UNIWARD with and without multiple sigmoid applications

Method	WOW	S-UNIWARD
	0.4 bpp	0.4 bpp
NO σ	87.45	86.85
Spatial σ	87.89	82.63
Channel σ	90.27	87.54
Proposed	90.77	88.69

Overall, the proposed method achieves the best performance under both the WOW and S-UNIWARD 0.4 bpp steganography algorithms. Moreover, the training converges faster and is more stable. This suggests that an adaptive approach, with sigmoid weight allocation before and after fusion, can more effectively enhance both spatial and channel performance, suppress irrelevant features, and highlight steganographic features. This, in turn, improves the model's ability to focus on key feature regions, thereby aiding in the recognition of steganographic signals.

Three sets of experiments are designed to compare the classification performance of network structures without the SPP module, with the traditional common pooling operation, and with the introduction of the SPP module under different steganography algorithms. The results in Table 10 show that after the

introduction of the SPP pyramid pooling module, the accuracy of the model on WOW and S-UNIWARD is improved to 90.77% and 88.69%, respectively, which is 1.82% and 2.59% higher than that without SPP, and significantly enhances the detection capability. In contrast, when the SPP module is not used, the accuracy drops to 88.95% and 86.10%; and when the ordinary pooling operation is used, the accuracy drops drastically to about 50%, and the training process shows serious convergence difficulties, which prevents the effective extraction of steganographic features. This result fully verifies the important role of the SPP module in improving the ability of multi-scale feature modeling and enhancing the model's perception of weak steganographic signals.

Table 10: Comparison of accuracy (%) without SPP, with normal pooling, and with SPP

Method	No SPP	Pool	SPP
WOW 0.4 bpp	88.95	50.15	90.77
S-UNIWARD 0.4 bpp	86.10	50.69	88.69

5 Conclusions

Most existing steganalysis models are convolution-based and mainly target grayscale images with high embedding payloads. However, these models often have difficulty in preserving subtle steganographic features, resulting in suboptimal detection accuracy and limited generalization capability. This study proposes a novel deep steganography analysis framework that integrates an adaptive attention mechanism with the lightweight DenseLite module to address these challenges. In the preprocessing stage, the DenseLite module is introduced to enhance the preservation of steganographic traces while minimizing the model complexity to provide a more efficient and effective feature base. For feature extraction, a dual-domain adaptive attention mechanism—combining channel and spatial attention—is designed to better highlight regions sensitive to steganographic modifications. In addition, a spatial pyramid pooling (SPP) module is employed to extract and fuse multi-scale features, which further improves the generalization capability.

With three mainstream steganography algorithms, the average accuracy of the proposed model is improved by 1.39% over the four payload levels. These results not only validate the effectiveness of combining lightweight design with attention mechanisms, but also provide empirical evidence that steganalysis can significantly benefit from domain-adaptive attention strategies and multiscale feature fusion. This study shows that, compared with the traditional method relying on deep convolutional stacking, combining the lightweight module with the attention mechanism can provide a more targeted and efficient solution for capturing weak cryptographic features, thus breaking through the bottleneck of the existing models that have limited detection capabilities in complex backgrounds. This study contributes to the creation of new knowledge by demonstrating how lightweight attention-enhancing architectures can effectively detect steganographic signals that were previously difficult to capture. The findings suggest new directions for building steganographic analysis models that are both efficient and highly generalizable, challenging the traditional reliance on heavily convoluted structures.

The findings of this study open up multiple paths for future research. Despite the encouraging results of the model in improving detection accuracy and generalization, further exploration is needed to enhance its adaptability to new and emerging steganographic methods. Future research can build on this model to explore more advanced feature extraction techniques or hybrid architectures that fuse traditional methods with deep learning-based approaches. In addition, extending the model's ability to handle more steganographic algorithms and image types will make it more versatile and robust in real-world scenarios. The proposed framework provides valuable insights into the design of efficient steganography analysis systems,

crucial for developing tools to counter emerging steganographic methods that current detection techniques cannot detect. Future research can build on this model to enhance robustness against noise and complex modifications, explore hybrid approaches combining traditional and deep learning methods, and improve detection systems' adaptability to new, unknown steganographic techniques.

Acknowledgement: We sincerely thank the editor and the anonymous reviewers for their careful review and valuable comments, which greatly improved the quality of this paper. We are grateful for the support from the Gansu Province Higher Education Institutions Industrial Support Program and the National Natural Science Foundation of China.

Funding Statement: This work was supported in part by Gansu Province Higher Education Institutions Industrial Support Program under Grant 2020C-29, and in part by the National Natural Science Foundation of China under Grant 61562002.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Zhenxiang He, Rulin Wu; data collection: Rulin Wu; analysis and interpretation of results: Rulin Wu, Xinyuan Wang; draft manuscript preparation: Xinyuan Wang, Rulin Wu; Zhenxiang He served as the corresponding author and oversaw data collection and analysis. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data will be made available by the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Luo W, Wei K, Li Q, Ye M, Tan S, Tang W, et al. A comprehensive survey of digital image steganography and steganalysis. APSIPA Trans Signal Inf Process. 2024;13(1). doi:10.1561/116.20240038.
2. Kheddar H, Hemis M, Himeur Y, Megías D, Amira A. Deep learning for steganalysis of diverse data types: a review of methods, taxonomy, challenges and future directions. Neurocomputing. 2024;581(3):127528. doi:10.1016/j.neucom.2024.127528.
3. Wei S, Wang Z, Zhang X. Universal image vaccine against steganography. Symmetry. 2025;17(1):66.
4. Shehab DA, Alhaddad MJ. Comprehensive survey of multimedia steganalysis: techniques, evaluations, and trends in future research. Symmetry. 2022;14(1):117. doi:10.3390/sym14010117.
5. Fridrich J, Goljan M, Hogea D. Steganalysis of JPEG images: breaking the F5 algorithm. Int Workshop Inf Hiding. 2002;2578(4):310–23. doi:10.1007/3-540-36415-3_20.
6. Harmsen JJ, Pearlman WA. Steganalysis of additive-noise modelable information hiding. Secur Watermarking Multimed Contents. 2003;5020:131–42. doi:10.1117/12.476813.
7. Li Z, Lu K, Zeng X, Pan X. Feature-based steganalysis for JPEG images. In: Proceedings of the 2009 International Conference on Digital Image Processing; 2009 Mar 7–9; Bangkok, Thailand.
8. Pevný T, Bas P, Fridrich J. Steganalysis by subtractive pixel adjacency matrix. In: Proceedings of the the 11th ACM Workshop on Multimedia and Security; 2009 Sep 7–8; New York, NY, USA.
9. Denemark T, Sedighi V, Holub V, Cogranne R, Fridrich J. Selection-channel-aware rich model for steganalysis of digital images. In: Proceedings of the 2014 IEEE International Workshop on Information Forensics and Security (WIFS); 2014 Dec 3–5; Atlanta, GA, USA.
10. Ma Y, Xu L, Zhang Y, Zhang T, Luo X. Steganalysis feature selection with multidimensional evaluation and dynamic threshold allocation. IEEE Trans Circuits Syst Video Technol. 2023;34(3):1954–69. doi:10.1109/tcsvt.2023.3295364.
11. Xu G, Wu H-Z, Shi Y-Q. Structural design of convolutional neural networks for steganalysis. IEEE Signal Process Lett. 2016;23(5):708–12. doi:10.1109/lsp.2016.2548421.

12. Fridrich J, Kodovsky J. Rich models for steganalysis of digital images. IEEE Trans Inf Forensics Secur. 2012;7(3):868–82. doi:10.1109/tifs.2012.2190402.
13. Ye J, Ni J, Yi Y. Deep learning hierarchical representations for image steganalysis. IEEE Trans Inf Forensics Secur. 2017;12(11):2545–57. doi:10.1109/tifs.2017.2710946.
14. Boroumand M, Chen M, Fridrich J. Deep residual network for steganalysis of digital images. IEEE Trans Inf Forensics Secur. 2018;14(5):1181–93. doi:10.1109/tifs.2018.2871749.
15. Yang J, Shi Y-Q, Wong EK, Kang X. JPEG steganalysis based on densenet. arXiv:1711.09335. 2017.
16. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the the IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21–26; Honolulu, HI, USA.
17. Zhang R, Zhu F, Liu J, Liu G. Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis. IEEE Trans Inf Forensics Secur. 2019;15:1138–50. doi:10.1109/tifs.2019.2936913.
18. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell. 2015;37(9):1904–16. doi:10.1109/tpami.2015.2389824.
19. Weng S, Sun S, Yu L. Fast SwT-based deep steganalysis network for arbitrary-sized images. IEEE Signal Process Lett. 2023;30(5):1782–6. doi:10.1109/lsp.2023.3336561.
20. He J, Weng S, Yu L, Chen D. Steganalysis network with two-branch preprocessing for spatial and JPEG domains. IEEE Trans Circuits Syst Video Technol. 2024;35(2):1451–63. doi:10.1109/tcsvt.2024.3470809.
21. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
22. Woo S, Park J, Lee J-Y, Kweon IS. Cbam: convolutional block attention module. In: Proceedings of the the European Conference on Computer Vision (ECCV); 2018 Sep 8–14; Munich, Germany.
23. Zhang X, Zhang X, Feng G. Image steganalysis network based on dual-attention mechanism. IEEE Signal Process Lett. 2023;30:1287–91. doi:10.1109/lsp.2023.3313517.
24. Ma Y, Wang J, Zhang X, Wang G, Xin X, Zhang Q. LGS-Net: a lightweight convolutional neural network based on global feature capture for spatial image steganalysis. IET Image Process. 2025;19(1):e70005. doi:10.1049/ipr2.70005.
25. Hu D, Zhou S, Shen Q, Zheng S, Zhao Z, Fan Y. Digital image steganalysis based on visual attention and deep reinforcement learning. IEEE Access. 2019;7:25924–35. doi:10.1109/access.2019.2900076.
26. Bravo-Ortiz MA, Mercado-Ruiz E, Villa-Pulgarin JP, Hormaza-Cardona CA, Quiñones-Arredondo S, Arteaga-Arteaga HB, et al. CVTStego-Net: a convolutional vision transformer architecture for spatial image steganalysis. J Inf Secur Appl. 2024;81(2):103695. doi:10.1016/j.jisa.2023.103695.
27. Guo F, Sun S, Weng S, Yu L, He J. A two-stream-network based steganalysis network: TSNet. Expert Syst Appl. 2024;255(5):124796. doi:10.1016/j.eswa.2024.124796.
28. Yedroudj M, Comby F, Chaumont M. Yedroudj-net: an efficient CNN for spatial steganalysis. In: Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2018 Apr 15–20; Calgary, AB, Canada.
29. Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21–26; Honolulu, HI, USA.
30. 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; 2016 Jun 26–30; Las Vegas, NV, USA.
31. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA.