文章编号:1674-2869(2025)05-0556-09

基于可逆网络结构的自适应鲁棒图像信息隐藏算法

易国洪,牛小青,张胜蓝
智能机器人湖北省重点实验室(武汉工程大学),武汉工程大学计算机科学与工程学院,
湖北 武汉 430205

摘 要:针对现有编码器-解码器结构的图像信息隐藏方法在安全性、嵌入后失真及抗噪声攻击能力上的不足,提出了一种基于可逆网络结构的自适应鲁棒图像信息隐藏算法。算法采用可逆隐藏块同时实现图像隐藏与恢复,通过集成轻量级卷积注意力模块,融合空间和通道注意力生成注意力权重,使网络根据权重自适应调整区域嵌入权重,在载体图像复杂纹理区域隐藏更多秘密信息。通过离散小波变换将信息嵌入至图像频域,通过目标损失指导在高频子带区域隐藏更多秘密信息。为进一步增强算法鲁棒性,添加了噪声层和信息增强模块,对含密图像施加高斯噪声和JPEG压缩,以模拟真实传输环境,并在恢复图像前利用信息增强模块对失真图像特征进行补偿。实验结果显示,所提出的算法在DIV2K数据集和COCO数据集上的平均峰值信噪比分别达到了44.82和41.03 dB,在应对传输环境高斯噪声和JPEG压缩情况下仍能有效恢复秘密图像,且在隐写分析残差网络(SRNet)中,检测准确率为55.01%,明显低于其他对比方法。该算法在保持视觉不可见性的同时增强了鲁棒性,为实际应用中的安全隐蔽通信提供了新的技术实现路径。

关键词:图像信息隐藏;可逆网络;注意力机制;鲁棒性
中图分类号:TP301.6 **文献标识码**:A **DOI**:10.19843/j.cnki.CN42-1779/TQ.202505017

Adaptive robust image information hiding algorithm based on an invertible network structure

YI Guohong, *NIU Xiaoqing*, *ZHANG Shenglan*
Hubei Key Laboratory of Intelligent Robot(Wuhan Institute of Technology);School of Computer Science & Engineering, Wuhan Institute of Technology, Wuhan 430205, China

Abstract: To address the limitations of existing encoder-decoder-based image information hiding methods in terms of security, post-embedding distortion, and resistance to noise attacks, an adaptive robust image information hiding algorithm based on an invertible network structure was proposed. The algorithm employs an invertible hidden block to simultaneously achieve image hiding and recovery. By integrating a lightweight convolutional attention module that combines spatial and channel attention, attention weights were generated, enabling the network to adaptively adjust regional embedding strength. This allowed more secret information to be hidden in complex texture regions of the carrier image. The information was embedded into the frequency domain of the image via discrete wavelet transform, with target loss guidance used to conceal more secret information in high-frequency subbands. To further enhance the robustness of the algorithm, a noise layer and an information enhancement module were incorporated. Gaussian noise and JPEG compression were applied to the secret-containing image to simulate real transmission conditions, while the information enhancement module compensated for distorted image features prior to recovery. Experimental

收稿日期:2025-05-23
基金项目:武汉工程大学研究生教育创新基金(CX2023321)
作者简介:易国洪,硕士,教授。Email: 178035026@qq.com
引文格式:易国洪,牛小青,张胜蓝. 基于可逆网络结构的自适应鲁棒图像信息隐藏算法[J]. 武汉工程大学学报,2025,47(5):556-564.

results showed that the proposed algorithm achieved average peak signal-to-noise ratios of 44.82 and 41.03 dB on the DIV2K dataset and COCO dataset, respectively. It effectively recovered secret images under the Gaussian noise and JPEG compression, and attained a detection accuracy of 55.01% with the steganalysis residual network (SRNet), which was significantly lower than that of other comparative methods. The algorithm enhances robustness while maintaining visual imperceptibility, offering a new technical path for secure covert communication in practical applications.

Keywords: image information hiding; invertible networks; attention mechanism; robustness

图像信息隐藏技术是一种将秘密信息不可感知地嵌入到数字图像中，必要时再将其从含密的图像中恢复并提取出来的技术[1]。根据嵌入方式的差别，可以将图像信息隐藏方法分为基于空域的方法和基于频域的方法。最低有效位(leatst significant bits，LSB)替换算法[2]基于修改像素LSB的原理隐藏秘密信息[3]，是一种经典的空域方法，但其大量像素位的修改可能会导致载体图像发生变化而泄露秘密信息的存在[4]。Khedmati等[5]将密码学与信息隐藏技术融合，通过结合2D混沌映射与离散小波变换实现信息隐藏，但隐藏容量较小。Jamal等[6]通过多次迭代将图像划分为不同片段，并利用LSB和最高有效位进行信息隐藏以提高安全性。Emad等[7]通过将秘密信息嵌入至整数小波变换逼近系数的LSB以完成信息隐藏。上述传统的图像信息隐藏方法隐藏安全性有限，很容易被高性能的隐写分析检测工具检测出秘密信息隐藏行为的存在，且隐藏容量十分有限。随着深度学习在图像处理任务中表现出的强大潜力，基于深度学习的隐藏方法逐渐成为信息隐藏的主要研究方向。Zhu等[8]提出了一种基于深度网络的数据隐藏(hiding data with deep networks，HiDDeN)方法用于图像隐藏。ur Rehman等[9]提出一种新的损失来约束编码器-解码器结构的网络，采用三通道RGB图像作为载体，实现了灰度图的隐藏。Zhang等[10]提出了一种基于生成对抗网络的高容量图像隐藏(high capacity image steganography with generative adversarial network，SteganoGAN)方法，用于将二进制信息隐藏到图像中，由编码器、解码器和判别器3个主要模块组成。Jing等[11]提出了一个基于可逆神经网络(invertible neural networks，INN)的深度图像隐藏(deep image hiding by INN，HiNet)框架，实现了将全尺寸的秘密图像隐藏到同规格载体图像中。Lu等[12]提出了一种可逆图像信息隐藏网络(invertible steganography network，ISN)，实现了大容量的图像信息隐藏。多图像隐藏于同一载体中是实现大容量信息隐藏的有效解决思路，Guan等[13]提出了一种基于INN的多图像信息隐藏框架(deep invertible network for multiple image hiding，DeepMIH)，聚焦于提升隐藏容量。Xu等[14]提出了一种基于流的鲁棒可逆图像信息隐藏(robust invertible image steganography，RIIS)框架，在应对含密图像失真情况下表现出了良好的性能。Yang等[15]基于HiNet提出了一种用于图像信息隐藏的实用且鲁棒的可逆网络(practical robust invertible network for image steganography，PRIS)，该网络聚焦于提升鲁棒性，并设计3步训练策略以提升性能。段新涛等[16]提出了一种基于卷积神经网络的轻量高效图像信息隐藏方法，采用尺度特征融合模块捕捉多维数据的复杂关系。刘连山等[17]提出了一种基于三通道融合的图像信息隐藏方法，通过主通道、底通道和中间通道融合提取特征以实现隐藏。上述基于深度学习的图像信息隐藏方法在进行信息隐藏时取得了良好的效果，但在实际的传输环境中，含密图像可能会经历不同类型的信道噪声攻击，接收方从得到的受损含密图像提取出秘密图像的准确度十分有限，秘密信息的提取质量难以得到保证。

针对现有图像信息隐藏方法存在的应对传输环境噪声鲁棒性不足、秘密信息嵌入可能导致载体图像失真明显的问题，提出了一种基于可逆神经网络[18]结构的自适应鲁棒图像信息隐藏算法(adaptive robust image information hiding algorithm based on invertible neural networks，ARINN)，应用融合通道和空间注意力的卷积注意力模块[19](convolutional block attention module，CBAM)增强卷积块的特征识别和提取能力，通过生成的注意力权重自适应指导在载体图像纹理复杂区域隐藏更多秘密信息，同时添加噪声训练并采用基于密集连接卷积块的信息增强模块对经受噪声攻击的含密图像进行特征增强，最大程度还原含密图像特征，提升算法的鲁棒性。

1 ARINN

1.1 ARINN模型

ARINN模型总体结构如图1所示。根据信息流向不同可以将模型分为前向隐藏过程和后向恢复过程，前向隐藏过程的输入是载体图像$\boldsymbol{X}_{\mathrm{c}}$和秘密图像$\boldsymbol{X}_{\mathrm{s}}$，首先经离散小波变换（discrete wavelet transform，DWT）操作将图像转换到频域，接着在一系列可逆隐藏块中完成秘密信息隐藏操作，最后再经过逆小波变换（inverse wavelet transform，IWT）输出得到含密图像$\boldsymbol{X}_{\mathrm{m}}$和丢失信息$\boldsymbol{r}$。为提升模型的鲁棒性，将包含不同强度高斯噪声和JPEG压缩的噪声层加入训练，并采用基于密集卷积网络[20]（dense convolutional network，DenseNet）架构的信息增强模块用于特征恢复，信息增强的含密图像$\boldsymbol{X}_{\mathrm{n}}$会先被送入信息增强模块以增强原始特征后，与辅助变量$\boldsymbol{z}$一同作为后向恢复过程的输入，经一系列可逆隐藏块以及IWT得到恢复载体图像$\boldsymbol{X}_{\mathrm{c,r}}$和恢复秘密图像$\boldsymbol{X}_{\mathrm{s,r}}$。

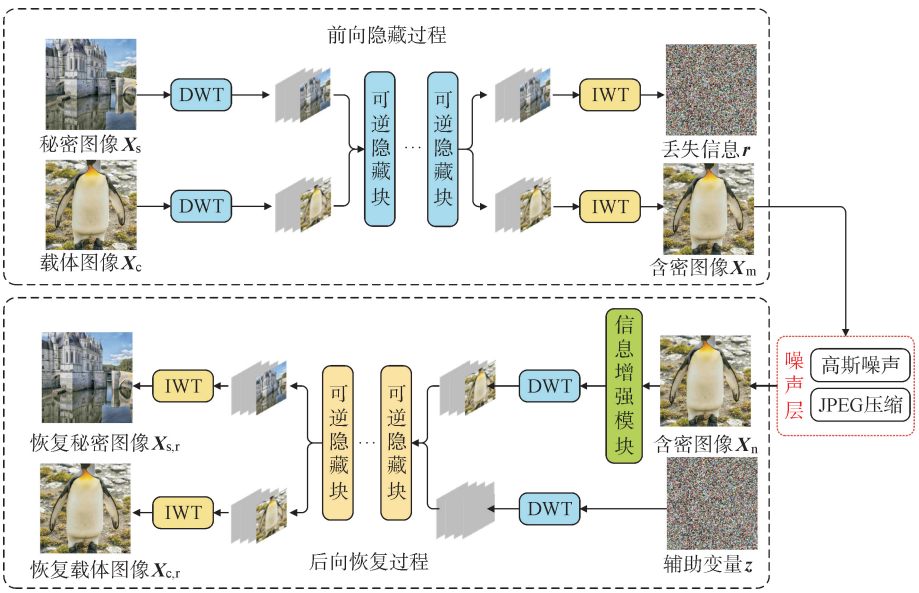

图1 ARINN模型

Fig. 1 ARINN model

在ARINN中，秘密图像被隐藏到载体图像的小波域而非像素域中，这是因为像素域的直接修改很容易导致载体图像出现伪影和颜色失真现象，而经过DWT后将图像分割成高频子带和低频子带，再进行信息隐藏，能够让网络更好地提取秘密图像和载体图像特征，从而达到更好的隐藏效果。

1.2 可逆隐藏块结构

可逆隐藏块是ARINN的核心部分，也是使网络保持可逆的关键，在ARINN中，前向隐藏过程和后向恢复过程各使用16个叠加的可逆隐藏块。对于前向隐藏过程中的第i个隐藏块，接收的输入为载体图像$\boldsymbol{X}_{\mathrm{c},i}$和秘密图像$\boldsymbol{X}_{\mathrm{s},i}$，通过式（1）和式（2）得到载体图像$\boldsymbol{X}_{\mathrm{c},i+1}$和秘密图像$\boldsymbol{X}_{\mathrm{s},i+1}$。其中，$\odot$表示两个向量逐元素相乘；$\alpha$是通过常数因子进行夹紧的Sigmoid函数；$\varphi$、$\eta$、$\rho$是3个具有相同结构的骨干网络，$\varphi$用于特征提取和转换，$\eta$对特征进行仿射变换，$\rho$用于对特征进行非线性变换。对输出的载体图像$\boldsymbol{X}_{\mathrm{c},i+1}$和秘密图像$\boldsymbol{X}_{\mathrm{s},i+1}$进行IWT，可得到含密图像$\boldsymbol{X}_{\mathrm{m}}$和丢失信息$\boldsymbol{r}$。

$$\boldsymbol{X}_{\mathrm{c},i+1}=\boldsymbol{X}_{\mathrm{c},i}+\varphi(X_{\mathrm{s},i}) \tag{1}$$

$$\boldsymbol{X}_{\mathrm{s},i+1}=\boldsymbol{X}_{\mathrm{s},i}\odot\exp(\alpha(\rho(\boldsymbol{X}_{\mathrm{c},i+1})))+\eta(\boldsymbol{X}_{\mathrm{c},i+1}) \tag{2}$$

后向提取过程中采用的可逆隐藏块结构相同，而数据流向相反。对于第i个可逆隐藏块，所接收的输入信息是由含密图像$\boldsymbol{X}_{\mathrm{m}}$和辅助变量$\boldsymbol{z}$经过DWT操作得到的$\boldsymbol{X}_{\mathrm{m},i+1}$和$\boldsymbol{z}_{i+1}$，输出得到$\boldsymbol{X}_{\mathrm{m},i}$和$\boldsymbol{z}_{i}$。其中，$\boldsymbol{z}$是服从高斯分布的随机采样。具体的转换过程见式（3）和式（4）。

$$\boldsymbol{z}_{i}=(\boldsymbol{z}_{i+1}-\eta(\boldsymbol{X}_{\mathrm{m},i+1}))\odot\exp(-\alpha(\rho(\boldsymbol{X}_{\mathrm{m},i+1}))) \tag{3}$$

$$\boldsymbol{X}_{\mathrm{m},i}=\boldsymbol{X}_{\mathrm{m},i+1}-\varphi(\boldsymbol{z}_{i}) \tag{4}$$

经一系列可逆隐藏块后转换输出$\boldsymbol{X}_{\mathrm{s},1}$，经过IWT得到恢复秘密图像$\boldsymbol{X}_{\mathrm{s,r}}$。图2为前向隐藏过程中的一个可逆隐藏块结构示例，其中DenseLayer代表密集卷积层。

现有图像信息隐藏方法一般只通过简单的卷积网络结构对载体图像和秘密图像进行特征提取，并不能精确识别到载体图像和秘密图像的多

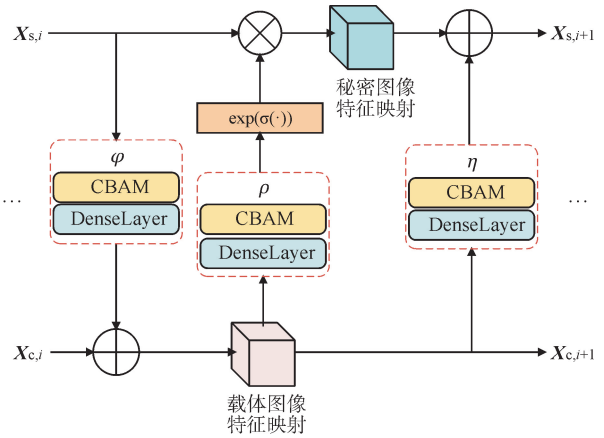

图 2　可逆隐藏块结构

Fig. 2　Reversible hidden block structure

尺度特征，导致方法不能有效应对隐写分析工具的检测。而注意力机制可以实现使网络只关注重要的特征，基于这一特性，ARINN 在可逆隐藏块中添加了通道注意力机制和空间注意力机制来充分学习输入图像的多尺度特征，具体实现方法是将 CBAM 集成到 φ、η、ρ 中，将 CBAM 的输入与 φ、η、ρ 原有的密集卷积层的输出相连，通过空间注意力，模型可以学习到图像中更适合隐藏信息的区域，根据注意力权重自适应调节区域嵌入权重，显著提升信息隐藏的隐蔽性。可逆隐藏块中集成 CBAM 的 φ、η、ρ 结构如图 3 所示，在可逆隐藏块中引入 CBAM 后依然保持网络的对称性，在前向隐藏过程中计算得到的注意力图在反向过程中都被精确保存和复用，不影响网络结构的理论可逆性。

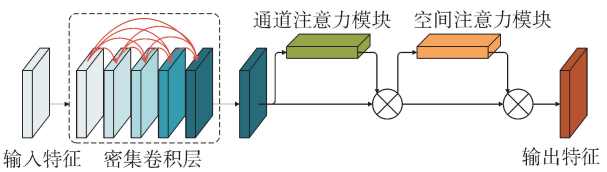

图 3　集成密集卷积层与 CBAM 的骨干网络结构

Fig. 3　Integrated dense convolutional layer and the backbone network structure of CBAM

1.3　信息增强模块结构

为增强模型的鲁棒性，ARINN 将噪声层加入训练用于模拟实际传输过程中图像可能遭受的攻击，对用于传输的含密图像施加两类可控强度的失真：高斯噪声（标准差 σ=1 和 σ=10），以及 JPEG 压缩，其中压缩质量因子（quality factor，QF）分别为 90 和 50。为了最大程度还原载体图像和秘密图像的特征，ARINN 添加了一个基于 DenseNet 结构的信息增强模块，取消了 DenseNet 原有的归一化层，接收受损的含密图像 $\boldsymbol{X}_{\mathrm{n}}$ 作为输入，用于增强和恢复含密图像原有特征。模块初始化 5 个卷积层，每个卷积层后伴随 LeakyReLU 激活函数以增加非线性。每个卷积层接收的输入都是前面所有卷积层输出的累积，这样每个层都可以接收到之前所有层的信息。第 1 层卷积处理原始输入，第 2 层卷积处理原始输入和第 1 层的输出，以此类推，直到最后 1 层卷积输出最终的特征映射。采用密集连接的思想，有助于网络捕捉更复杂的特征，以提升含密图像在抵抗噪声和其他失真时的稳定性。

1.4　ARINN 算法

ARINN 采用基于可逆网络结构的隐藏与提取机制，首先，分别对载体图像和秘密图像执行 Haar 离散小波变换，在保留全局结构信息的同时将细节与纹理部分分离，之后利用多个可逆隐藏块嵌入秘密信息。拼接后的特征依次通过 N_1 个可逆隐藏块，在每个隐藏块内部，首先使用密集卷积层提取秘密分支特征，并施加通道-空间双重注意力指导嵌入，通道维度上通过全局均值与最大池化获得权重，空间维度再利用卷积得到权重，突出更为鲁棒的嵌入位置。隐藏块内部结合密集卷积层与 CBAM，分别计算特征映射、缩放和偏移参数，从而更新载体与秘密特征，完成信息嵌入。所有可逆块迭代完成后，网络从最终特征张量中按既定顺序取回更新后的 4 个载体子带，并把剩余通道分离为丢失信息 $\boldsymbol{r}$。随后对更新后的子带执行 IWT 得到含密图像 $\boldsymbol{X}_{\mathrm{m}}$。具体步骤如下：

（1）对载体图像 $\boldsymbol{X}_{\mathrm{c}}$、秘密图像 $\boldsymbol{X}_{\mathrm{s}}$ 各进行 1 次 DWT，将二者分解为低频子带 LL 及 3 个高频子带 LH、HL、HH 后，再将两组子带串联为统一特征张量 $\boldsymbol{F}_0$。

（2）循环采用 N_1=16 个可逆隐藏块，在第 k 层处理时，输入特征 $\boldsymbol{F}_{k-1}$ 由两部分组成：$\boldsymbol{F}_{k-1}=[\boldsymbol{X}_{\mathrm{c,f}},\boldsymbol{X}_{\mathrm{s,f}}]$，其中 $\boldsymbol{X}_{\mathrm{c,f}}$ 为载体分支特征，$\boldsymbol{X}_{\mathrm{s,f}}$ 为秘密分支特征。按照式（5）进行纹理信息提取，提炼局部细纹理与高频细节信息，得到经过密集卷积层特征信息提取操作的秘密分支特征 $\boldsymbol{X}_{\mathrm{s,dense}}$。

$$\boldsymbol{X}_{\mathrm{s,dense}}=D(\boldsymbol{X}_{\mathrm{s}}) \tag{5}$$

其中，$D(\cdot)$ 代表密集卷积层特征信息提取操作。

（3）通过施加通道注意力机制，对 $\boldsymbol{X}_{\mathrm{s,dense}}$ 分别执行全局平均池化与最大池化操作，并将二者结果求和，经两层 1×1 卷积与 Sigmoid 得到通道注意力权重 $\boldsymbol{M}_{\mathrm{c}}$，再施加空间注意力对经 $\boldsymbol{M}_{\mathrm{c}}$ 加权后的特征进行全局平均池化和最大池化并沿通道拼接，取 7×7 卷积后经 Sigmoid 得到空间注意力权重 $\boldsymbol{M}_{\mathrm{s}}$。根据式（6）计算得到区域嵌入权重 $\varphi(\boldsymbol{X}_{\mathrm{s,f}})$，按照相

同的步骤可以分别计算得到缩放因子$\rho(\boldsymbol{X}_{\mathrm{c,f}})$和偏移量$\eta(\boldsymbol{X}_{\mathrm{c,f}})$。

$$\varphi(\boldsymbol{X}_{\mathrm{s,f}})=\boldsymbol{X}_{\mathrm{s,f}}\odot\boldsymbol{M}_{\mathrm{c}}\odot\boldsymbol{M}_{\mathrm{s}} \tag{6}$$

(4)在可逆隐藏块中进行前向信息隐藏映射，按照式(7)定义进行载体加法注入。根据式(8)定义进行秘密信息映射调整，指数exp保证尺度因子始终为正，从而避免因符号反转可能导致的信息丢失，整个更新过程对$\boldsymbol{X}_{\mathrm{s}}$构成仿射变换，其对数行列式仅依赖$\rho$和$\alpha$。

$$\boldsymbol{X}_{\mathrm{c,f}}=\boldsymbol{X}_{\mathrm{c,f}}+\varphi(\boldsymbol{X}_{\mathrm{s,f}}) \tag{7}$$

$$\boldsymbol{X}_{\mathrm{s,f}}=\boldsymbol{X}_{\mathrm{s,f}}\odot\exp\left[\alpha(\rho(\boldsymbol{X}_{\mathrm{c,f}}))\right]+\eta(\boldsymbol{X}_{\mathrm{c,f}}) \tag{8}$$

(5)将计算得到的特征$\boldsymbol{X}_{\mathrm{c,f}}$和$\boldsymbol{X}_{\mathrm{s,f}}$拼接送入下一层可逆隐藏块中，循环结束得到最终特征$\boldsymbol{F}_{16}$。

(6)按照初始顺序提取更新后的LL、LH、HL、HH子带，剩余通道信息标记为丢失信息$\boldsymbol{r}$，之后执行IWT合成得到含密图像$\boldsymbol{X}_{\mathrm{m}}$并输出。

输出含密图像$\boldsymbol{X}_{\mathrm{m}}$后，前向隐藏阶段结束，算法输出视觉上与原始载体图像$\boldsymbol{X}_{\mathrm{c}}$几乎无差异却携带秘密的$\boldsymbol{X}_{\mathrm{m}}$以及辅助恢复所需的$\boldsymbol{r}$。在解码端，只需将$\boldsymbol{X}_{\mathrm{m}}$与$\boldsymbol{r}$按相反方向通过相同的可逆隐藏块序列，即可提取出恢复秘密图像。

1.5 损失函数

在损失函数设计时，ARINN考虑前向隐藏损失、后向恢复损失、信息增强损失、低频小波损失4种损失函数的加权值作为整个模型的损失。在前向隐藏过程中，ARINN关注原始载体图像和嵌入秘密信息后的含密图像之间的差异，前向隐藏损失函数值可由式(9)求出，其中，$\|\cdot\|_2$表示向量的2-范数，θ代表网络的参数，N是整个训练样本集的数量，$\boldsymbol{X}_{\mathrm{c},n}$代表第$n$张载体图像，$\boldsymbol{X}_{\mathrm{m},n}$代表第$n$张含密图像。

$$L_{\mathrm{conceal}}(\theta)=\sum_{n=1}^{N}\left\|\boldsymbol{X}_{\mathrm{c},n}-\boldsymbol{X}_{\mathrm{m},n}\right\|_2 \tag{9}$$

在后向恢复过程中，ARINN关注原始秘密图像和恢复秘密图像之间的差异，后向恢复损失函数值由式(10)计算得出，式中，$\boldsymbol{X}_{\mathrm{s},n}$代表第$n$张原始秘密图像，$\boldsymbol{X}_{\mathrm{s,r},n}$代表第$n$张恢复秘密图像，$E_z$代表对$z$取数学期望。同时，网络使用服从高斯分布的辅助变量$z$来提取秘密图像，以便提升提取后秘密图像的准确性。

$$L_{\mathrm{reveal}}(\theta)=\sum_{n=1}^{N}E_z\left(\left\|\boldsymbol{X}_{\mathrm{s},n}-\boldsymbol{X}_{\mathrm{s,r},n}\right\|_2\right) \tag{10}$$

信息增强模块的目的是对受损含密图像进行特征补偿和恢复，信息增强损失主要关注信息增强前的第n张含密图像$\boldsymbol{X}_{\mathrm{n},n}$和信息增强后的第$n$张含密图像$\boldsymbol{X}_{\mathrm{n,e},n}$之间的差异，损失函数值由式(11)计算得出。

$$L_{\mathrm{enhance}}(\theta)=\sum_{n=1}^{N}\left\|\boldsymbol{X}_{\mathrm{n},n}-\boldsymbol{X}_{\mathrm{n,e},n}\right\|_2 \tag{11}$$

为了确保更多的信息能够隐藏在高频子带中，ARINN同时关注前向隐藏过程中生成的含密图像的低频子带和经DWT的原始载体图像的低频子带之间的差异。低频小波损失函数值由式(12)计算得出，其中，$H(\boldsymbol{X}_{\mathrm{c},n})_{\mathrm{L}}$代表经Haar小波变换后的原始载体图像的低频子带，$H(\boldsymbol{X}_{\mathrm{m},n})_{\mathrm{L}}$代表含密图像的低频子带。

$$L_{\mathrm{frequency}}(\theta)=\sum_{n=1}^{N}\left(\left\|H(\boldsymbol{X}_{\mathrm{c},n})_{\mathrm{L}}-H(\boldsymbol{X}_{\mathrm{m},n})_{\mathrm{L}}\right\|_2\right) \tag{12}$$

模型的总体损失由以上4种损失函数值加权计算得出，具体的计算公式见式(13)，其中，λ_1、λ_2、λ_3和λ_4是平衡不同损失项的权重，模型训练过程中，首先将λ_4设置为0，以最小化前向隐藏损失、后向恢复损失和信息增强损失的方式对网络进行预训练，之后再添加低频小波损失，保证整个模型依旧端到端训练。

$$L_{\mathrm{total}}=\lambda_1 L_{\mathrm{conceal}}+\lambda_2 L_{\mathrm{reveal}}+\lambda_3 L_{\mathrm{enhance}}+\lambda_4 L_{\mathrm{frequency}} \tag{13}$$

2 结果与分析

2.1 条件设置

实验采用DIV2K[21]训练数据集来训练ARINN，并分别采用DIV2K测试数据集、COCO[22]数据集来测试模型的性能，实验环境采用GeForce RTX 4090，Python3.7，PyTorch1.7。为了让模型有更好的泛化能力，首先将训练图像进行预处理，按照随机裁剪策略裁剪为224×224像素，而测试图像则采用中心裁剪策略统一处理。可逆隐藏块的数量设置为16，采用三阶段训练策略训练ARINN，每个阶段训练1 600个迭代，第一阶段初始学习率设置为$10^{-4.5}$，之后将学习率逐步减小至$10^{-5.0}$、$10^{-5.2}$。从训练数据集中随机选择1/2作为载体图像样本，另1/2为秘密图像样本，采用Adam[23]优化器标准参数优化训练，λ_1、λ_2、λ_3和λ_4初始分别设置为1.0、0.1、0.1和1.0，通过计算出训练图像和测试图像的平均峰值信噪比(peak signal-to-noise ratio，PSNR)和平均结构相似性系数(structural similarity index measure，SSIM)定量评估ARINN的表现。为了对比ARINN与其他主流方法的表现差异，在GeForce RTX 4090环境下，对HiDDeN[8]、SteganoGAN[10]、HiNet[11]、PRIS[15]

等4种方法重新进行训练，并展开定量评估。各方法的模型参数量与训练策略如下：HiDDeN[8]的参数量为11.0×10⁶，采用Adam优化器（学习率为10⁻⁴），训练迭代次数为5 000；SteganoGAN[10]的参数量为28.3×10⁶，使用Adam优化器（学习率为2×10⁻⁴），迭代5 000次；HiNet[11]的参数量为17.7×10⁶，采用Adam优化器（学习率为10⁻⁴），迭代5 000次；PRIS[15]的参数量为16.2×10⁶，采用Adam优化器（学习率为10⁻⁴），迭代4 500次；而ARINN的参数量为16.5×10⁶，同样采用Adam优化器（学习率为10⁻⁴），训练迭代4 800次。

2.2　隐藏与提取质量分析

图4展示了采用不同方法进行信息隐藏的隐藏视觉质量对比。由图4可知，采用ARINN隐藏秘密图像时，含密图像与载体图像的残差几乎不可见，表明其信息隐藏过程对载体图像的影响极小。相比之下，采用SteganoGAN和HiDDeN进行信息隐藏后的残差图仍然能看到部分秘密图像的细节。

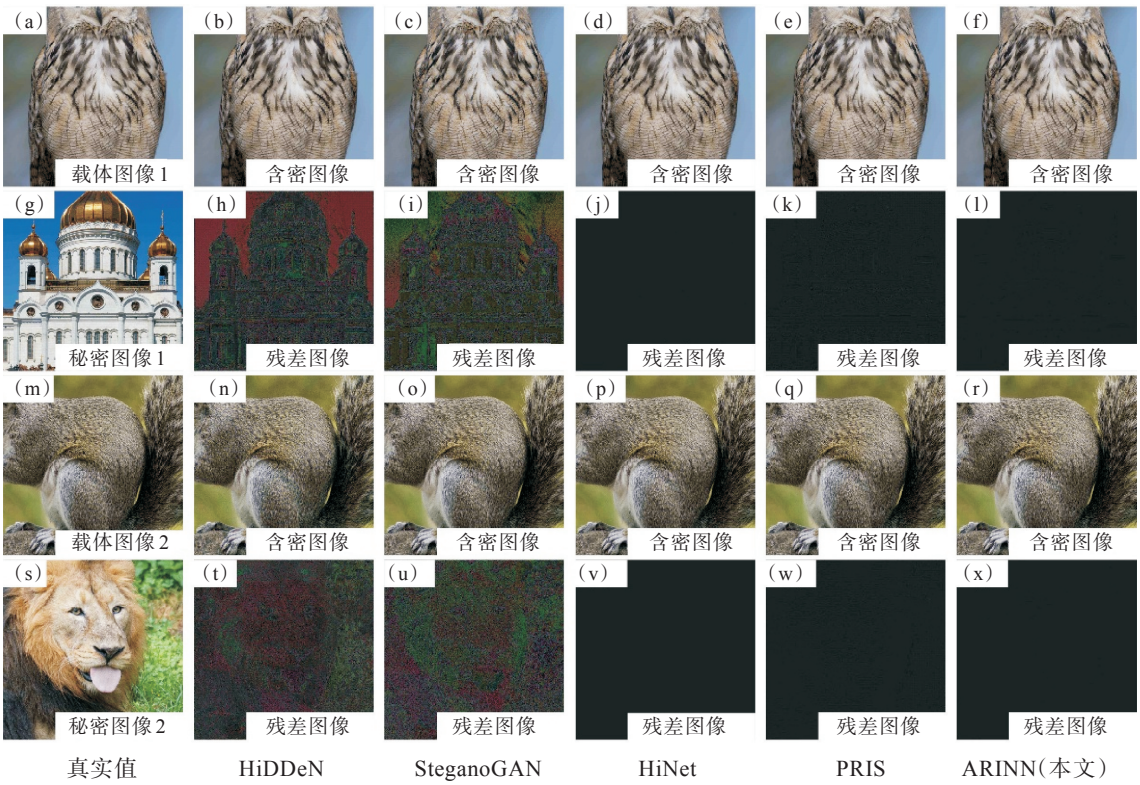

图4　不同方法隐藏视觉质量对比

Fig. 4　Comparison of the visual quality of different methods of hiding

为定量评估ARINN的隐藏与提取性能，使用相同数据集在GeForce RTX 4090环境下训练HiDDeN[8]、SteganoGAN[10]、HiNet[11]、PRIS[15]等主流图像信息隐藏模型，并分别在DIV2K测试数据集和COCO测试数据集上进行统一测试，计算载体图像和含密图像、秘密图像和恢复秘密图像两对图像之间的平均PSNR和平均SSIM，并将两者作为性能比较指标。其中，在DIV2K测试数据集上的载体图像和含密图像的平均PSNR和平均SSIM结果如表1所示，可以看出，在DIV2K数据集中ARINN的平均PSNR达到了44.82 dB，在所有对比方法中表现最好，这表明ARINN的含密图像与原始载体图像之间的失真较小，视觉质量高。相较于SteganoGAN，ARINN的平均PSNR提升了8.91 dB，表明其在高质量图像上具有显著的优势。在COCO数据集中，平均PSNR为41.03 dB，基本与HiNet持平。

表1　不同数据集载体图像和含密图像平均PSNR和平均SSIM

Table 1　Average PSNR and average SSIM of carrier images and encrypted images in different datasets

方法	平均PSNR / dB		平均SSIM	
	DIV2K	COCO	DIV2K	COCO
HiDDeN[8]	34.64	35.03	0.949	0.952
SteganoGAN[10]	35.91	37.43	0.955	0.965
HiNet[11]	43.63	41.24	0.984	0.976
PRIS[15]	38.69	39.01	0.969	0.971
ARINN	44.82	41.03	0.987	0.980

2.3　鲁棒性分析

为测试ARINN的鲁棒性，分别将高斯噪声标

准差σ设置为1和10,以及使用不同强度的JPEG压缩(QF=50和QF=90)对含密图像加扰,并评估ARINN恢复秘密图像的质量。表2展示了在DIV2K数据集上的实验结果,表3展示了在COCO数据集上的表现。可以看出,未考虑鲁棒性设计的SteganoGAN和HiNet在噪声下难以恢复秘密图像,平均PSNR显著降低。在DIV2K数据集上,高斯噪声标准差σ=1时,ARINN的平均PSNR达到了29.64 dB,显示出ARINN在低强度噪声下具有较好的图像质量恢复能力,当σ增至10时,ARINN的平均PSNR依然保持在28.12 dB,均高于HiDDeN的20.64 dB和SteganoGAN的6.21 dB,表明在高强度噪声下ARINN仍能够显著减小图像质量损失。在JPEG压缩实验中,ARINN在QF=50条件下的平均PSNR值为27.91 dB,相较于PRIS的25.03 dB提升了2.88 dB,表明ARINN在低质量压缩情况下的抗压缩性能更优。在QF=90压缩条件下,ARINN的平均PSNR为30.02 dB,在所有方法中仍表现最佳。在COCO数据集上的实验结果进一步验证了ARINN的鲁棒性,在应对标准差σ=10的高强度高斯噪声时ARINN的平均PSNR达到28.93 dB,显著优于HiNet的8.22 dB,验证了ARINN的泛化能力。

表2 DIV2K数据集中秘密图像和恢复秘密图像平均PSNR

Table 2 Average PSNR of secret images and restored secret images in DIV2K dataset

方法	平均PSNR / dB				
	σ=1	σ=10	QF=50	QF=90	无噪声
HiDDeN[8]	23.82	20.64	22.91	25.24	33.82
SteganoGAN[10]	8.69	6.21	7.34	9.44	34.55
HiNet[11]	12.05	9.13	11.82	13.24	48.74
PRIS[15]	28.98	26.46	25.03	28.03	36.63
ARINN	29.64	28.12	27.91	30.02	48.04

表3 COCO数据集中秘密图像和恢复秘密图像平均PSNR

Table 3 Average PSNR of secret images and recovered secret images in COCO dataset

方法	平均PSNR / dB				
	σ=1	σ=10	QF=50	QF=90	无噪声
HiDDeN[8]	25.02	23.64	21.98	24.73	35.03
SteganoGAN[10]	9.03	8.21	8.34	9.44	36.55
HiNet[11]	10.14	8.22	12.62	13.92	47.43
PRIS[15]	27.92	27.39	26.11	29.44	37.05
ARINN	30.06	28.93	29.10	31.12	47.28

图5展示了ARINN在不同噪声和压缩强度下恢复秘密图像的质量对比,可以看出,在无噪声和高斯噪声σ=1的低强度噪声情况下,残差图几乎为黑色,说明信息隐藏过程对图像的影响极小。在高斯噪声σ=10时,虽然残差图中开始出现少量细节,但总体结构和细节依然清晰,表明ARINN在高强度噪声下依旧能够保持较高的恢复质量。对于JPEG压缩,随着压缩质量的降低(QF=90和QF=50),恢复图像的细节有所减少,但主要特征仍然保留。特别是在低质量的QF=50压缩条件下,残差图显示出少量的压缩伪影,但恢复的图像依然保持了大部分视觉信息。这说明ARINN在应对高压缩率的JPEG攻击时依然具备良好的鲁棒性。

2.4 抗隐写检测分析

隐写分析技术可以判断输入的图像中是否包含有秘密信息,隐写分析的结果是衡量隐藏方法安全性的重要评价指标。为了验证ARINN的安全性,分别采用基于统计学的隐写分析方法和基于深度学习的隐写分析方法对ARINN进行评估。StegExpose是一款经典的基于统计学的隐写分析工具,从COCO数据集中随机选择1 000张图像用于测试,使用ARINN进行图像信息隐藏,并完成秘密图像提取,使用StegExpose工具在大范围内改变检测阈值获得受试者工作特征(receiver operating characteristic, ROC)曲线,结果如图6所示,ARINN的ROC曲线下面积为0.509,十分接近于0.500,说明从统计学角度分析很难区分出原载体图像和含密图像的区别。

此外,采用基于深度学习的隐写分析残差网络[24](steganalysis residual network, SRNet)评估ARINN,SRNet是可用于隐写分析的网络,可有效检测出载体图像和含密图像的差异。实验结果表明,ARINN在SRNet检测下的准确率为(55.01±0.23)%,明显低于HiDDeN[(81.04±0.20)%]、SteganoGAN[(99.64±0.11)%]、HiNet[(56.32±0.03)%]和PRIS[(58.18±0.09)%]等对比方法。上述结果表明,即使在面对深度学习隐写分析模型时,ARINN仍能保持良好的隐蔽性。

2.5 模型效率分析

ARINN采用基于可逆隐藏块的可逆结构网络进行信息隐藏与提取,在同等通道数的前提下,可逆隐藏块与常规残差块的卷积算子数量一致。为分析模型效率,分别用乘加运算量(floating-point operations, FLOPs)和单张推理帧率(frames per

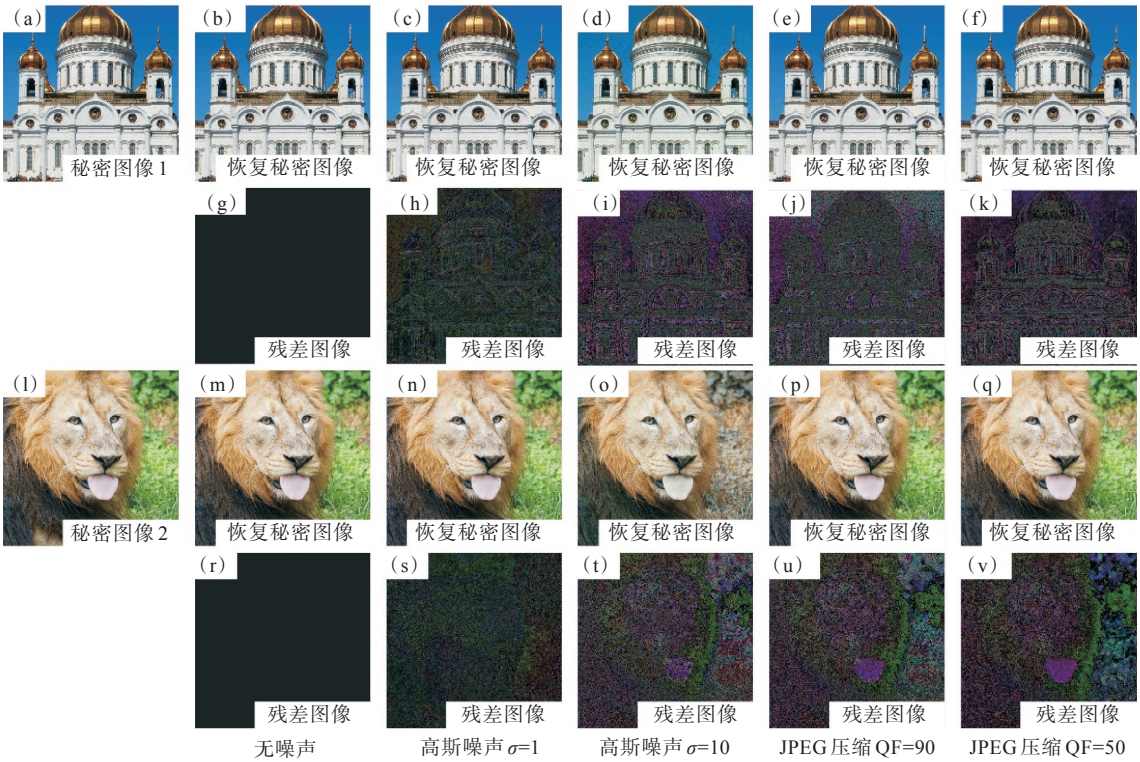

图5　不同噪声环境下恢复秘密图像质量对比

Fig. 5　Quality comparison of recovered secret images under different noise environments

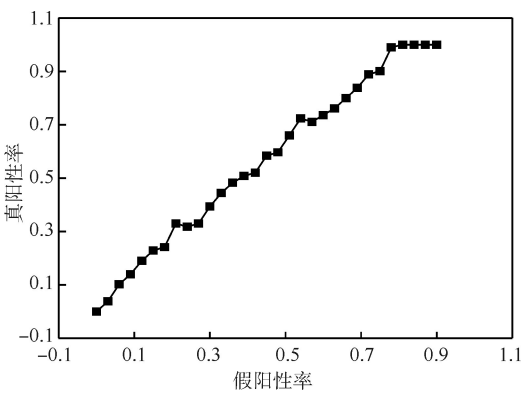

图6　使用StegExpose检测工具得到的ROC曲线

Fig. 6　ROC curves obtained using the StegExpose inspection tool

second，FPS）两个客观指标衡量各模型的计算效率，其中，FLOPs代表每张图像前向推理需要的乘加运算次数，FPS代表模型每秒可处理224×224像素图像的数量。表4展示了本文方法与其他方法的模型效率对比结果，所有模型均在GeForce RTX 4090、PyTorch1.7环境下运行并计算，图像输入尺寸均为224×224像素。可以看出，采用可逆网络结构的框架HiNet、PRIS、ARINN的计算量与常规残差块相近，仅比HiDDeN略高，但远低于包含大规模GAN生成器的SteganoGAN。HiNet的FPS值为48.8，在所有对比模型中最优，而ARINN将CBAM嵌入耦合分支，FPS值为46.0，比HiNet慢约6%，仍满足实时隐藏和提取的需求。

表4　与其他方法的模型效率对比

Table 4　Comparison of modeling efficiency with other methods

方法	可逆网络结构	FLOPs / 10^9	FPS
HiDDeN[8]	×	17.9	46.2
SteganoGAN[10]	×	41.4	34.1
HiNet[11]	√	19.6	48.8
PRIS[15]	√	21.1	45.3
ARINN	√	20.7	46.0

3　结　论

针对现有图像信息隐藏方法存在的隐藏隐蔽性不足、抗噪鲁棒性较差的问题，提出了一种基于可逆网络结构的自适应鲁棒图像信息隐藏算法ARINN，算法基于可逆网络结构实现，通过添加融合空间和通道注意力机制的CBAM模块，增强了网络对复杂纹理区域的特征提取能力，使隐藏过程更加隐蔽和自适应。同时，ARINN添加了噪声层，并采用基于密集连接卷积网络的信息增强模块在后向恢复前补偿含密图像的细节，从而确保恢复图像的质量。实验结果表明，ARINN在平均PSNR和平均SSIM上的表现优于多数现有方法，生成的含密图像具有高度的不可见性和抗隐写分

析能力。即使在高强度噪声和压缩条件下，ARINN仍能够有效地恢复秘密信息。

参考文献

[1] 张卫明，王宏霞，李斌，等. 多媒体隐写研究进展[J]. 中国图象图形学报，2022，27(6)：1918-1943.

[2] CHAN C K，CHENG L M. Hiding data in images by simple LSB substitution [J]. Pattern Recognition，2004，37(3)：469-474.

[3] TAMIMI A A，ABDALLA A M，AL-ALLAF O. Hiding an image inside another image using variable-rate steganography [J]. International Journal of Advanced Computer Science and Applications，2013，4(10)：18-21.

[4] GUPTA S，GUJRAL G，AGGARWAL N. Enhanced least significant bit algorithm for image steganography [J]. International Journal of Computational Engineering & Management，2012，15(4)：40-42.

[5] KHEDMATI Y，PARVAZ R，BEHROO Y. 2D Hybrid chaos map for image security transform based on framelet and cellular automata [J]. Information Sciences，2020，512：855-879.

[6] JAMAL S S，FARWA S，ALKHALDI A H，et al. A robust steganographic technique based on improved chaotic-range systems[J]. Chinese Journal of Physics，2019，61：301-309.

[7] EMAD E，SAFEY A，REFAAT A，et al. A secure image steganography algorithm based on least significant bit and integer wavelet transform [J]. Journal of Systems Engineering and Electronics，2018，29(3)：639-649.

[8] ZHU J R，KAPLAN R，JOHNSON J，et al. HiDDeN：hiding data with deep networks[C]// Lecture Notes in Computer Science. Berlin：Springer，2018：682-697.

[9] ur REHMAN A，RAHIM R，NADEEM S. End-to-end trained CNN encoder-decoder networks for image steganography [C]// Lecture Notes in Computer Science. Berlin：Springer，2019：723-729.

[10] ZHANG K A，CUESTA-INFANTE A，XU L，et al. SteganoGAN：high capacity image steganography with GANs [Z/OL]. (2019-01-30) [2025-06-27]. https://arxiv.org/abs/1901.03892.

[11] JING J P，DENG X，XU M，et al. HiNet：deep image hiding by invertible network[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway：IEEE，2021：4713-4722.

[12] LU S P，WANG R，ZHONG T，et al. Large-capacity image steganography based on invertible neural networks [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway：IEEE，2021：10811-10820.

[13] GUAN Z Y，JING J P，DENG X，et al. DeepMIH：deep invertible network for multiple image hiding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2023，45(1)：372-390.

[14] XU Y M，MOU C，HU Y J，et al. Robust invertible image steganography[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway：IEEE，2022：7865-7874.

[15] YANG H，XU Y T，LIU X H，et al. PRIS：practical robust invertible network for image steganography [J]. Engineering Applications of Artificial Intelligence，2024，133：108419.

[16] 段新涛，白鹿伟，徐凯欧，等. 基于卷积神经网络的轻量高效图像隐写[J]. 应用科学学报，2025，43(1)：80-93.

[17] 刘连山，黄瑜. 基于三通道深度融合技术的图像隐写方法[J]. 信息安全研究，2025，11(3)：257-264.

[18] DINH L，KRUEGER D，BENGIO Y. NICE：non-linear independent components estimation [Z/OL]. (2015-04-10) [2025-06-27]. https://arxiv. org/abs/1410.8516.

[19] WOO S，PARK J，LEE J Y，et al. CBAM：convolutional block attention module [C]// Lecture Notes in Computer Science. Berlin：Springer，2018：3-19.

[20] HUANG G，LIU Z，PLEISS G，et al. Convolutional networks with dense connectivity [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2022，44(12)：8704-8716.

[21] AGUSTSSON E，TIMOFTE R. NTIRE 2017 challenge on single image super-resolution：dataset and study[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway：IEEE，2017：1122-1131.

[22] LIN T Y，MAIRE M，BELONGIE S，et al. Microsoft COCO：common objects in context [C]//Lecture Notes in Computer Science. Berlin：Springer，2014：740-755.

[23] KINGMA D P，BA J L. Adam：a method for stochastic optimization [Z/OL]. (2017-01-22) [2025-06-27]. https://arxiv.org/abs/1412.6980.

[24] BOROUMAND M，CHEN M，FRIDRICH J. Deep residual network for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security，2019，14(5)：1181-1193.

本文编辑：霍淑利