软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2021,32(9):2901−2915 [doi: 10.13328/j.cnki.jos.005980]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于卷积神经网络的低嵌入率空域隐写分析*

沈 军[1,2], 廖 鑫[1,3], 秦 拯[1,2], 刘绪崇[3]

[1](湖南大学 信息科学与工程学院,湖南 长沙 410012)

[2](大数据研究与应用湖南省重点实验室(湖南大学),湖南 长沙 410082)

[3](网络犯罪侦查湖南省重点实验室(湖南警察学院),湖南 长沙 410138)

通讯作者: 廖鑫, E-mail: xinliao@hnu.edu.cn

摘 要: 近年来,基于深度学习的空域隐写分析研究在高嵌入率下已经取得了较好的成果,但是对低嵌入率的检测效果还不太理想.因此设计了一种卷积神经网络结构,使用 SRM 滤波器进行预处理来获取隐写噪声残差,采用 3 个卷积层并对卷积核大小进行合理设计,通过适当选择批量归一化操作和激活函数来提升网络的性能.实验结果表明:与现有方法相比,所提出的网络结构对 WOW,S-UNIWARD 和 HILL 这 3 种常见的空域内容自适应隐写算法取得了更好的检测效果,且在低嵌入率 0.2bpp,0.1bpp 和 0.05bpp 下的检测效果有非常明显的提升.还提出了逐步迁移(step by step)的迁移学习方法,进一步提升低嵌入率条件下的隐写分析效果.

关键词: 隐写分析;卷积神经网络;低嵌入率;迁移学习

中图法分类号: TP309

中文引用格式: 沈军,廖鑫,秦拯,刘绪崇.基于卷积神经网络的低嵌入率空域隐写分析.软件学报,2021,32(9):2901−2915. http://www.jos.org.cn/1000-9825/5980.htm

英文引用格式: Shen J, Liao X, Qin Z, Liu XC. Spatial steganalysis of low embedding rate based on convolutional neural network. Ruan Jian Xue Bao/Journal of Software, 2021,32(9):2901−2915 (in Chinese). http://www.jos.org.cn/1000-9825/5980.htm

Spatial Steganalysis of Low Embedding Rate Based on Convolutional Neural Network

SHEN Jun[1,2], LIAO Xin[1,3], QIN Zheng[1,2], LIU Xu-Chong[3]

[1](College of Computer Science and Electronic Engineering, Hunan University, Changsha 410012.China)

[2](Hunan Key Laboratory of Big Data Research and Application (Hunan University), Changsha 410082, China)

[3](Hunan Key Laboratory of Cybercrime Reconnaissance (Hunan Police Academy), Changsha 410138, China)

Abstract: In recent years, the research of spatial steganalysis based on deep learning has achieved sound results under high embedding rate, but the detection performance under low embedding rate is still not ideal. Therefore, a convolutional neural network structure is proposed, which uses the SRM filter for preprocessing to obtain implicit noise residuals, adopts three convolution layers and designs the size of convolution kernel reasonably, and selects appropriate batch normalization operations and activation functions to improve the network performance. The experimental results show that compared with the existing methods, the proposed network can achieve better detection performance for WOW, S-UNIWARD, and HILL, three common adaptive steganographic algorithms in spatial domain, and

* 基金项目: 国家自然科学基金(61972142, 61402162, 61772191); 湖南省自然科学基金(2017JJ3040); 模式识别国家重点实验室开放课题(201900017); 湖南省科技计划(2015TP1004, 2016JC2012); 网络犯罪侦查湖南省普通高校重点实验室开放课题(2017WLFZZC001)

Foundation item: National Natural Science Foundation of China (61972142, 61402162, 61772191); Hunan Provincial Natural Science Foundation of China (2017JJ3040); Open Project Program of the National Laboratory of Pattern Recognition (201900017); Science and Technology Key Projects of Hunan Province (2015TP1004, 2016JC2012); Open Research Fund of Key Laboratory of Network Crime Investigation of Hunan Provincial Colleges (2017WLFZZC001)

收稿时间: 2019-06-18; 修改时间: 2019-09-23; 采用时间: 2019-11-06; jos 在线出版时间: 2020-04-21

significant improvement in detection performance at low embedding rates of 0.2 bpp, 0.1 bpp, and 0.05 bpp. A step-by-step transfer learning method is also designed to further improve the steganalysis effect under low embedding rate conditions.

Key words: steganalysis; convolution neural network; low embedding rate; transfer learning

隐写的主要原理是将秘密信息隐藏在原始载体的不易被人感知的冗余信息中,从而达到通过载体传递秘密信息而不被察觉的目的[1−3].隐写分析的主要目的是检测载体中是否隐藏了秘密信息,目前,常见的载体包括图像、文本、音频、视频等多媒体信息[4].随着隐写技术的不断进步,近几年提出的 WOW[5],S-UNIWARD[6]和 HILL[7]等空域自适应隐写算法能够自动将秘密信息隐藏在图像纹理较为复杂的区域,使得图像能够保持很复杂的统计特征.传统的空域隐写分析方法为了应对这些更复杂的隐写技术,需要将特征设计得更加复杂且维度更高,如空间丰富模型 SRM[8].但是传统的空域隐写分析方法需要人为地设计特征,隐写分析的效果取决于特征的设计.并且随着隐写算法的发展,特征的设计也变得越来越复杂和困难,同时,也极大地延长了训练时间.近几年,随着深度学习的发展并不断在计算机视觉领域取得研究成果,研究人员也开始将深度学习的方法应用到了空域隐写分析领域中,并取得了较好的研究成果[9].

虽然目前已有很多基于卷积神经网络的隐写分析研究取得了一定的进展,但是这些研究都是在高嵌入率的条件下进行检测,在低嵌入率的情况下进行检测的效果还有待进一步提高.特别是在 0.1bpp 甚至是 0.05bpp 这种低嵌入率下的检测效果,目前的隐写分析方法都很难正确地进行检测.因此,本文针对低嵌入率设计了一个新的网络结构进行空域隐写分析,网络使用 SRM 滤波器进行预处理操作,并结合使用两个小尺寸卷积核和一个大尺寸卷积核,使网络能够有效地捕获低嵌入率的隐写特征.网络中通过取消池化层,防止隐写特征信息的丢失.网络的浅层使用 TanH 激活函数,深层使用 ReLU 激活函数,并通过批量归一化操作对网络性能进行进一步提升.实验结果表明:本文设计的网络结构在对 WOW[5],S-UNIWARD[6]和 HILL[7]这 3 种常见的空域内容自适应隐写算法进行隐写分析时,检测效果与现有基于卷积神经网络的隐写分析方法相比取得了明显的提升.在对低嵌入率(0.2bpp,0.1bpp 和 0.05bpp)进行检测时,本文提出的网络结构能够得到比较理想的检测效果.为了实现对低嵌入率的检测效果进行进一步提升,本文还提出了一种逐步迁移(step by step)的迁移学习方法.

本文第 1 节首先介绍相关工作.第 2 节主要介绍本文设计的卷积神经网络结构,详细介绍了网络的预处理层、卷积层、批量归一化和激活函数,并对模型特点进行了分析,最后通过迁移学习提高检测效果.第 3 节对实验参数、实验设置和实验结果进行介绍.第 4 节对全文进行总结.

1 相关工作

目前,基于卷积神经网络的隐写分析方法,网络结构的设计以两个卷积层、三个卷积层或五个卷积层为主,本文根据卷积层的深度分别对现有方法进行介绍.

1.1 两层卷积结构

Pibre 等人在文献[10]中针对使用相同密钥生成隐写图像这一场景,设计了 Pibre-Net.Pibre-Net 的预处理使用单个 KV 核的高通滤波器层(HPF 层),网络结构中只有两个卷积层,且分别使用 7×7 和 5×5 的大尺寸卷积核.Pibre-Net 中去掉了池化层,直接通过卷积层来减小特征图维度,避免池化操作造成隐写特征信息的丢失.在该场景下,对 Bossbase 图像库进行 S-UNIWARD[6]嵌入率为 0.4bpp 的检测,Pibre-Net 的准确性相比 SRM 有了很大的提升.基于 Pibre[10]提出的场景,Salomon 等人[11]直接使用大尺寸卷积核设计了 Salomon-Net 网络结构,Salomon-Net 的输入图像大小为 512×512,该网络只有两个卷积层:第 1 个卷积层作为全局过滤器输出一个特征图;第 2 个卷积层使用了 509×509 的大尺寸卷积核,输出 64 个 2×2 大小的特征图.Salomon-Net 中对 WOW[5]和 HUGO[12]分别进行了检测,实验结果显示:该模型不仅在 0.4bpp 嵌入率下能够取得很好的检测结果,而且在 0.1bpp 低嵌入率下的检测效果也比较理想.

高培贤等人在文献[13]中设计了两层卷积层和两层全连接层的浅层网络结构 S-CNN,该结构同样使用高通滤波器层(HPF 层)作为预处理操作.与 Xu-Net[14]相比,S-CNN 减少了卷积的层数,同时,通过去除池化层来避

免隐写噪声信息的丢失.该文献在实验中使用Bossbase图像库对S-UNIWARD[6]算法0.4bpp嵌入率进行了检测,检测效果优于Tan-Net[15],Qian-Net[16]和Xu-Net[14].

1.2 三层卷积结构

Tan等人在文献[15]中首次将深度学习的方法应用于空域隐写分析领域中,构造了包含3个卷积层和一个全连接层的4层网络结构,使用了KV核对第1层卷积核进行初始化,通过利用卷积自动编码器(SCAE)进行预训练,检测效果有了较大提升.在Tan-Net的工作中,验证了随机初始化第1个卷积层的训练模型基本没有隐写分析检测的能力.

1.3 五层卷积结构

Qian-Net[16]是由Qian等人提出的一个具有5个卷积层的网络结构,使用KV核作为预处理层对图像进行预处理,使得模型能够直接对残差图像进行学习,降低了图像内容对训练的干扰.Qian-Net还根据隐写噪声的特点,使用了高斯激活函数和均值池化,进一步提高了检测性能.Qian-Net在Bossbase图像库中的检测准确性相比SRM[8]只低了3%~5%,在ImageNet图像中的检测准确性与SRM相当.Qian-Net在基于深度学习的隐写分析中属于很好的研究成果.Qian等人提出采用迁移学习[17]的方法提高模型在低嵌入率下的检测性能,将高隐写容量的训练模型迁移到低隐写容量中进行训练,该方案在减少训练时间的同时,有效地提高了检测正确率.

Xu等人随后提出了Xu-Net[14],该网络使用KV核作为高通滤波器层(HPF层)对图像进行预处理操作,网络中使用5个卷积层,第1个卷积层之后,利用绝对(ABS)层来消除残差信号的符号影响.前两个卷积层的卷积核为5×5,为了防止网络模型过拟合,在随后的卷积层中使用1×1大小的卷积核.每个卷积层中使用了批量归一化(batch normalization,简称BN)操作,前两个卷积层后使用TanH激活函数,其他卷积层使用ReLU激活函数.每个卷积层通过均值池化来减小特征图的维度,均值池化能够综合所有残差信息,降低信息丢失的影响.Xu-Net在Bossbass图像库中,对S-UNIWARD[6]和HILL[7]算法的检测能力均优于SRM.Xu等人在文献[18]中对之前的工作进行了改进,使用卷积神经网络的集成学习和重叠池化方法来提高检测效果.

Yedroudj等人在文献[19]中通过结合Xu-Net[14]和Res-Net[20]网络的特点设计了Yedroudj-Net,该网络预处理层使用了30个SRM[8]卷积核,让网络能够提取到更多的隐写特征.网络结构中使用5个卷积层,综合使用了绝对(ABS)层、批量归一化(batch normalization,简称BN)层、截断函数(truncation function,简称Trunc)[21]和均值池化层.该文献中使用Bossbase图像库分别对WOW[5],S-UNIWARD[6]算法进行检测,发现0.4bpp和0.2bpp嵌入率下的效果均优于SRM[8],Xu-Net[14]和Ye-Net[22].

1.4 其他深层结构

基于Xu等人的研究[14],Ye等人提出了Ye-Net[22],该网络使用了更深的八层卷积网络结构,并且使用30个SRM[8]卷积核作为预处理层来让模型学习更多的特征.Ye等人在文献中设计了新的截断线性单元(truncated linear unit,简称TLU)作为激活函数,通过适当的设置参数T(一般取3或7),使网络能够更好地适应隐写噪声分布.Ye-Net还通过选择通道,进一步提高了该模型的检测效果.实验中,该文献结合Bossbase和BOWS2这两个图像库进行检测,其准确率已显著优于SRM等传统隐写分析方法.Wu等人利用残差网络Res-Net[20]构造了一个深层次的隐写分析模型Wu-Net[23],Wu-Net通过增加卷积层数量,使模型能够更有效的捕获图像的隐写特征.Wu-Net的检测效果均优于SRM算法[8]、Qian-Net[16]和Xu-Net[12].Tsang等人为了使模型能够对任意尺寸图像进行处理,基于Ye-Net[22]网络结构提出了Tsang-Net[24].Tsang-Net中,在全连接层前加入了统计矩提取层,统计矩提取层通过将卷积层输出的任意大小特征图转换为固定维度的特征输入全连接层.Tsang-Net实现了对任意大小的图像进行隐写分析检测,且保持了较好的检测能力.

2 本文提出的卷积神经网络结构

针对低嵌入率下空域隐写分析存在的问题,本文构建了一个新的卷积神经网络结构Shen-Net实现空域隐写分析[25].Shen-Net整体算法框架如图1所示,主要分为输入模块、卷积模块和输出模块.待测图像首先进入预

处理层对待测图像进行预处理操作,预处理操作能够从待测图像中提取出噪声残差信息,有利于卷积模块的特征学习;提取的噪声残差信息随后进入卷积模块,卷积模块中,首先通过卷积运算提取隐写特征,紧随的批量归一化操作和激活层能够有效提高卷积模块的特征学习能力和提升网络的性能,合理设计多组卷积模块能够使网络能够更好地学习隐写特征;经过卷积模块中一系列的卷积层和激活层等层层连接之后,需要通过全连接层进行连接,并将全连接层输出值直接传给分类器 Softmax 层进行分类,最终输出的分类结果为待测图像分属原始图像和携密图像的概率值.为了有效提升低嵌入率下的隐写分析效果,本文基于 Shen-Net 框架提出了逐步迁移学习方案,将高嵌入率下的训练模型作为初始参数逐步迁移至低嵌入率的网络中进行训练,使低嵌入率的网络能够有效借助高嵌入率训练模型的参数作为辅助来提升对低嵌入率隐写特征的学习能力.

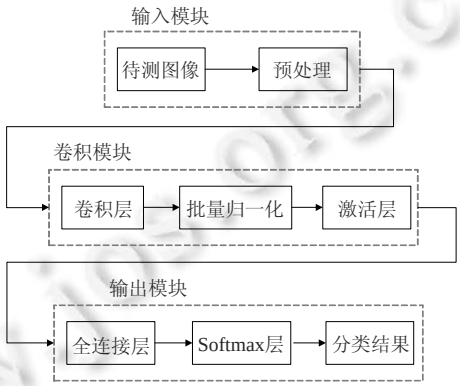


Fig.1 Overall framework of the algorithm proposed in this paper

图 1 本文提出的算法整体框架

网络的整体结构如图 2 所示.Shen-Net 的输入图像的大小为 256×256,网络结构包括一个预处理层、3 个卷积模块.其中,每个卷积模块包括卷积层、批量归一化操作、激活函数,卷积模块后跟随了两个全连接层,最后使用 softmax 函数进行分类.下面将对 Shen-Net 网络结构的各个部分进行详细阐述.

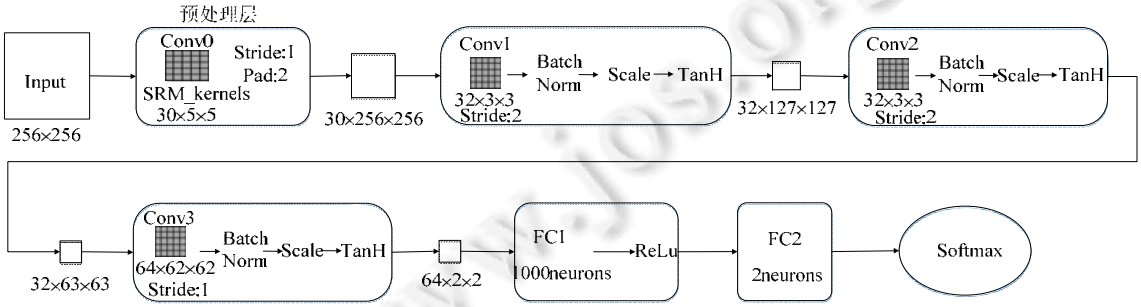


Fig.2 Shen-Net convolutional neural network structure

图 2 Shen-Net 卷积神经网络结构

2.1 网络执行流程

本文提出的 Shen-Net 执行流程框图如图 3 所示,在模型的训练阶段,训练图像输入网络后首先通过预处理层进行噪声残差提取,在后续网络的传播中对噪声残差图像进行隐写特征的学习训练.训练过程中,利用正向传播计算输出结果,通过对输出结果与实际结果求偏差,判断是否超过容许范围;否则进行反向传播,并计算各层中的误差,并通过梯度下降算法更新各层权值.通过反复传播计算,最终生成训练好的网络模型.

在测试阶段,直接将待检测图像输入网络模型中,模型同样首先进行噪声残差的提取操作,再通过后续的传播计算,最终进行分类,分类结果为两个标签上的概率值.最后,根据图像分类产生的概率值进行最终判定,概率

值大的做为最终结果.

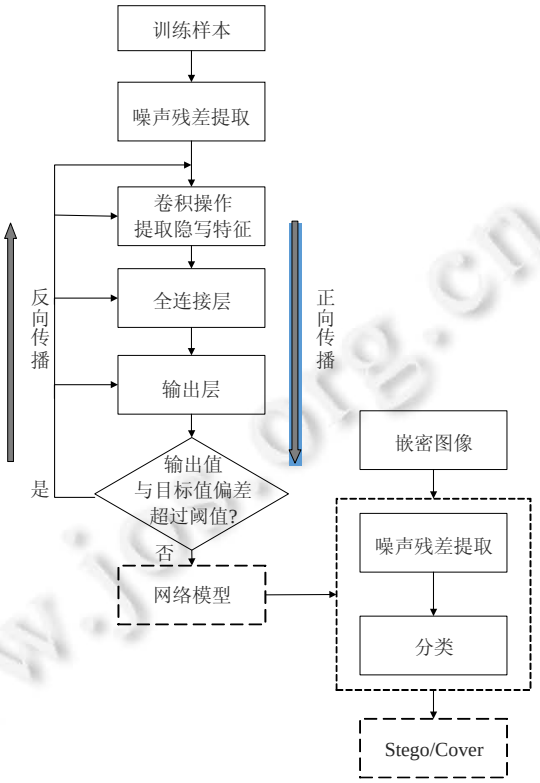


Fig.3 Shen-Net execution flowchart

图 3 Shen-Net 执行流程框图

2.2 预处理层

预处理层主要用于提取出输入图像的噪声残差分量,因为信息隐藏操作可以被视为向载体图像添加极低幅度的噪声[22].若将图像直接输入卷积层,很难保证卷积操作能够有效地提取出隐写噪声,从而导致模型的收敛速度会非常慢[10,16].而预处理操作的主要目的是为了增强隐写信号和图像信号之间的信噪比,并抑制图像内容对训练过程所造成的影响,因此,对输入图像进行噪声残差提取的预处理操作,能够有效提升模型对隐写特征的学习效果.图像噪声残差的计算公式如下所示:

$$R=X*K \tag{1}$$

其中,*表示卷积操作;X 表示输入图像;K 表示用于计算噪声残差的线性滤波器,其目的是为了通过相邻元素值对中心元素值进行预测,并计算差值.因此,噪声残差的计算可以通过卷积层进行模拟[19],本文采用的预处理操作主要基于 Fridrich 等人[8]的研究,预处理层使用了一个具有 30 个滤波器的卷积层,卷积核大小为 5×5,卷积核的初始权值使用空间丰富模型(SRM)中的 30 个高通滤波核进行初始化[19,22].其中,预处层中所使用的几类高通滤波核分别如下所示:

$$1st=[-1\ +1] \tag{2}$$

$$2nd=[+1\ -2\ +1] \tag{3}$$

$$3rd=[+1\ -3\ +3\ -1] \tag{4}$$

$$EDGE\ 3\times3=\begin{bmatrix}-1 & +2 & -1\\ +2 & -4 & +2\end{bmatrix} \tag{5}$$

$$SQUARE\ 3\times3=\begin{bmatrix}-1 & +2 & -1\\ +2 & -4 & +2\\ -1 & +2 & -1\end{bmatrix} \tag{6}$$

$$EDGE\ 5\times5=\begin{bmatrix}-1 & +2 & -2 & +2 & -1\\ +2 & -6 & +8 & -6 & +2\\ -2 & +8 & -12 & +8 & -2\end{bmatrix} \tag{7}$$

$$SQUARE\ 5\times5=\begin{bmatrix}-1 & +2 & -1 & +2 & -1\\ +2 & -6 & +8 & -6 & +2\\ -2 & +8 & -12 & +8 & -2\\ +2 & -6 & +8 & -6 & +2\\ -1 & +2 & -1 & +2 & -1\end{bmatrix} \tag{8}$$

其中,“1st”“2nd”和“3rd”经过 45 度旋转操作后,可以分别得到 8 个、4 个和 8 个滤波核;“*EDGE* 3×3”和“*EDGE* 5×5”进行 90 度旋转操作后,分别可以得到 4 个滤波核.为了将 30 个滤波核尺寸统一为 5×5,在“1st”“2nd”“3rd”和“*EDGE* 3×3”的四周填充 0.相比于只使用一个滤波核的高通滤波器层(HPF 层)进行预处理操作[10,13,14,16],30 个 SRM 滤波核组合了 7 种不同的滤波残差模型,从而 SRM 滤波器进行预处理操作时能够更有效地提取出隐写图像中的噪声残差分量,从而有利于后续卷积操作进行特征提取和学习,加快模型在训练过程中的收敛速度.

2.3 卷积层

第一个卷积模块中,卷积层使用 32 个滤波器,卷积核大小为 3×3.卷积层的输入为预处理进行噪声残差分量提取后的噪声残差图,从第一个卷积层开始提取噪声残差中的隐写特征,并生成用于下一阶段计算的特征图.卷积操作的计算公式如下所示:

$$X_j^l=\sum_i X_i^{l-1}*K_{ij}^l+B_j^l \tag{9}$$

其中,*表示卷积计算,X_j^l 表示第 l 层的第 j 张特征图,X_j^{l-1} 表示第 l–1 层的第 i 张特征图,K_{ij}^l 表示用来连接第 l 层的第 i 张输入特征图和第 j 张输出特征图的卷积核,B_j^l 表示第 l 层的第 j 张特征图的偏置量.

由于池化层是一个下采样的过程,在减小特征图大小的同时,会使得部分隐写特征信息丢失,从而降低后续卷积操作进行特征提取的性能,收敛速度变得缓慢,从而影响模型最终的分类准确率.因此,在本文的网络结构中取消了池化层的使用.但是为了减小卷积层的特征图大小,同时降低卷积操作的计算量,本文通过设置卷积层中卷积核的大小和步长来完成.因此,本文第 1 个卷积层的步长设为 2,经过第 1 个卷积层的卷积操作后的特征图大小为 32×127×127.

Shen-Net 中的第 2 个卷积层设置与第 1 个卷积层相同,也是使用了 32 个卷积核为 3×3 大小的滤波器.本文通过多组实验验证了与第一层使用同样数量卷积核,训练模型能够得到更好的检查效果,实验结果见表 1.检测算法为 S-UNIWARD,隐写强度为 0.1bpp.实验中,对 3 个卷积层设置 3 组不同的卷积核数量,实验结果分别给出了隐写图像(stego)、原始图像(cover)的检测正确率和平均检测正确率,本文根据实验结果选取了最优的参数设置.表 1 中最后一组实验取消了第 2 个卷积层,在第 1 个卷积层后直接进入大卷积核进行卷积操作,实验结果表明,其他使用 3 个卷积层的模型检测性能更好.Shen-Net 中进行第 2 个卷积操作后,输出特征图的大小为 32×63×63.

Table 1 S-UNIWARD 0.1bpp detection accuracy of different convolution kernel settings

表 1 S-UNIWARD 算法 0.1bpp 不同卷积核设置的检测准确率

Conv1	Conv2	Conv3	Stego (%)	Cover (%)	Average (%)
32×3×3	32×3×3	64×62×62	80.88	85.47	83.18
64×3×3	32×3×3	16×62×62	79.94	84.27	82.11
64×3×3	32×3×3	64×62×62	80.23	84.38	82.31
32×3×3	*null*	64×126×126	79.87	84.07	81.97

在第 3 个卷积模块的卷积层中使用了 64 个滤波器,特别的是使用了大小为 62×62 的大尺寸卷积核.Salomon 等人[11]验证了使用大卷积核能够构建小的长程相关模式,可以获得一组精简的识别特征.通过小卷积核与大卷积核的结合使用,使得模型在训练阶段能够有效地学习到低嵌入率的隐写特征.大卷积核还能够保证网络正确学习隐写特征的同时,降低特征的维度.第三个卷积层输出的特征图大小为62×2×2,极大地减小了特征图的尺寸,减轻了后续的计算复杂度.

2.4 批量归一化

通过使用批量归一化层(BN 层)[21]对每个卷积层实现归一化操作.由于卷积神经网络中的训练阶段,每一层都会对网络参数进行更新,而每一层对参数的更新都会影响后续网络输入数据的分布,并随着网络深度的加深进行放大,输入数据的分布的变化会降低对网络训练的收敛速度.批量归一化操作能够很好地解决数据分布产生变化的问题,归一化后的值都在特定的范围以内,使得模型能够快速地进行收敛,并在一定程度上防止网络出现过拟合的现象.批量归一化首先需对特征的每个维度进行归一化,其公式如下所示:

$$\hat{x}^{(k)}=\frac{x^{(k)}-E[x^{(k)}]}{\sqrt{Var[x^{(k)}]}} \tag{10}$$

其中,$E[x^{(k)}]$表示每一批训练数据 $x^{(k)}$的平均值,$Var[x^{(k)}]$表示每一批训练数据方差,$\sqrt{Var[x^{(k)}]}$表示数据的标准差.通过公式(10)变化之后的数据形成了均值为 0、方差为 1 的正态分布.为了避免以上变换影响网络的特征学习能力,第 2 步需要进行变换重构,其公式如下所示:

$$y^{(k)}=\gamma^{(k)}\hat{x}^{(k)}+\beta^{(k)} \tag{11}$$

其中,γ和β为可学习的参数,每个训练数据 $x^{(k)}$都有对应的γ、β参数;$y^{(k)}$为批量归一化操作后输出的特征图.当$\gamma=\sqrt{Var[x^{(k)}]}$且$\beta=E[x^{(k)}]$时,可以得到上一层学到的特征分布.通过变换重构,使得网络对特征分布的学习没有被破坏.网络中,通过BN层实现归一化,将卷积层输出的数据进行归一化处理;Scale层实现归一化后的平移和缩放.通过批量归一化操作,能够避免梯度消失,并能有效提高模型的检测精度[26].

2.5 激活函数

本文在前 3 个卷积模块的最后使用了 TanH 函数作为激活函数,TanH 函数如公式(12)所示:

$$f(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{12}$$

第 1 个全连接层后使用 ReLU 函数作为激活函数,ReLU 函数如公式(13)所示:

$$f(x)=\max(0,x) \tag{13}$$

其中,x 为输入特征图.TanH 函数能够使输出与输入的关系保持非线性单调上升和下降关系.Xu 等人[14]验证了,在网络结构的前几组非线性激活函数使用 TanH 函数能够比 ReLU 函数取得更好的性能.由于 TanH 函数的饱和区域,能够有效地限制数据值的范围.

经过 3 个卷积模块的特征提取后进入分类模块,分类模块主要包括两个全连接层和一个损失函数:第 1 个全连接层的神经元数量为 1 000 个,其后使用 ReLU 激活函数来提升网络的性能;第 2 个全连接层的神经元数量为 2,对应于网络输出的类别.分类模块的最后,通过 softmax 函数来对两个类别标签上产生概率分布.

2.6 模型分析

本文在预处理阶段使用了 30 个 SRM 滤波器,相比于仅仅使用一个 HPF 高通滤波器的预处理方式,SRM 滤波器能够提取出更多的噪声残差信息,使网络在训练阶段学习到更多的隐写特征,从而提高隐写分析能力.在 Pibre[10]、Salomon[11]等人的网络结构中,都是只采用了两个卷积层,本文增加了卷积层的层数,有利于训练阶段进行特征提取.

通过在第 3 个卷积层中使用大尺寸卷积核,在提高了特征提取性能的同时,降低了特征数量,从而限制了训练模型的大小(在输入图像大小为 256×256 时,文献[10]模型大小约为 980MB,本文的训练模型大小约为 31MB).

且通过结合前两个卷积层的小尺寸卷积核与第 3 个卷积层的大尺寸卷积核,使网络能够有效地捕获低嵌入率下的隐写噪声特征.

本文在卷积层中取消了池化层的使用,避免了池化层的下采样操作造成隐写特征的丢失.基于 Xu 等人[14]的研究,在 3 个卷积层后使用 TanH 函数作为激活函数,第 1 个全连接层后使用 ReLU 函数作为激活函数,通过在浅层 TanH 函数和深层 ReLU 函数的结合使用,一定程度上提升了网络性能.

Shen-Net 网络能够对低嵌入率的隐写特征进行有效学习,并使模型最终进行正确分类,关键在于预处理层、小尺寸卷积核和大尺寸卷积核的结合使用.预处理层中,通过 30 个高通滤波器,从低嵌入率隐写图像中提取出 30 种不同的噪声残差图像.在卷积模块中,通过小尺寸卷积核与大尺寸卷积核对噪声残差图像的卷积操作,使模型能够有效地提取低嵌入率下微弱的隐写特征.通过对网络的层数进行一定控制,让模型的复杂度尽量缩小.同时,Shen-Net 中,通过批量归一化操作来控制各卷积模块中的数据分布,并在网络的卷积层中,通过 TanH 函数进行非线性激活操作,提高网络的表达能力,使网络收敛性得到有效保证.由于低嵌入率时图像中的隐写特征信息本就非常微弱,而池化层进行下采样操作时无可避免地会对特征图的信息产生丢失,因此在 Shen-Net 中取消了池化层,避免了池化操作所造成的隐写特征的丢失,间接地提高了模型的检测能力.通过对网络结构的合理设计,使 Shen-Net 的模型能够对低嵌入率的隐写图像的检测能够达到很好的效果.

2.7 逐步迁移(step by step)的迁移学习

由于内容自适应隐写算法[5−7]会根据图像纹理特征和嵌入率,将信息从纹理最复杂的区域开始嵌入,因此当嵌入率很低时,秘密信息会被嵌入在图像最复杂的纹理区域中,从而导致网络在训练阶段难以学习到足够的隐写特征,使训练模型的检测效果并不理想.

本文为了进一步提升低嵌入率的检测效果,在相同的隐写算法中,通过迁移学习的方法将高嵌入率下的训练模型作为初始参数迁移至低嵌入率的网络中进行训练,使低嵌入率的网络能够借助高嵌入率训练模型的参数作为辅助,来提升对隐写特征的学习能力.文献[17]中,利用迁移学习的方法有效地提升了 Qian-Net[16]的检测能力.迁移学习框架如图 4 所示,在本文的迁移学习方法中,使用的网络结构为 Shen-Net 结构,源任务与目标任务使用相同的网络结构.首先,通过对高嵌入率的图像集进行训练,利用高嵌入率下训练好的模型,将训练参数迁移至嵌入率较低的目标任务中对网络训练参数进行初始化.目标任务在训练阶段通过对网络参数进行微调来提升低嵌入率下训练模型的隐写分析能力.通过利用高嵌入率中训练参数迁移来初始化低嵌入率的训练任务,来代替使用随机值对参数进行初始化,能够有效提升网络的训练效果.

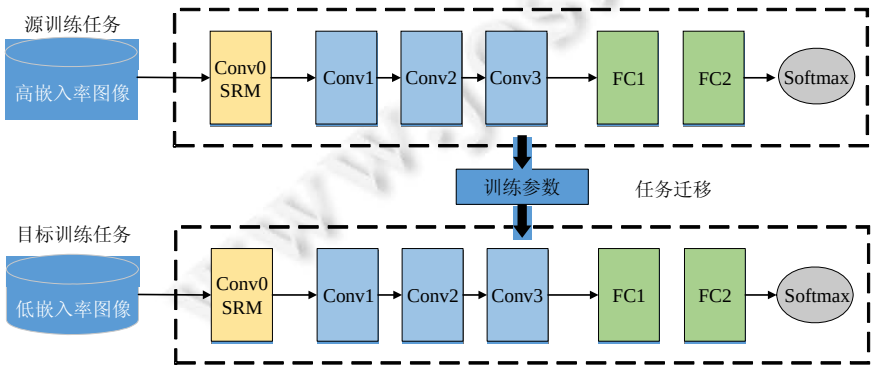


Fig.4 Shen-Net embedding rate transfer learning framework

图 4 Shen-Net 嵌入率迁移学习框架

嵌入率越高的训练模型,往往能够更好地学习到隐写特征,因此,本文希望能够将最高嵌入率下的训练模型的参数直接迁移至低嵌入率下进行微调,但如果源任务与目标任务之间的嵌入率相差过大(如 0.5bpp 与 0.05bpp 之间),隐写噪声信息存在强度上的差异,直接进行迁移学习可能不会取得很好的性能提升.因此,本文提出了逐

步迁移(step by step)的方法,即:将 0.5bpp 的训练模型迁移至 0.4bpp 中进行训练,再将 0.4bpp 的训练模型迁移至 0.3bpp,以此类推,直至 0.05bpp.通过逐步迁移的方式,将最高嵌入率下的训练参数逐步迁移至低嵌入率下进行微调,消除了噪声信息的强度差异.

3 实验结果与分析

3.1 数据集和实验平台

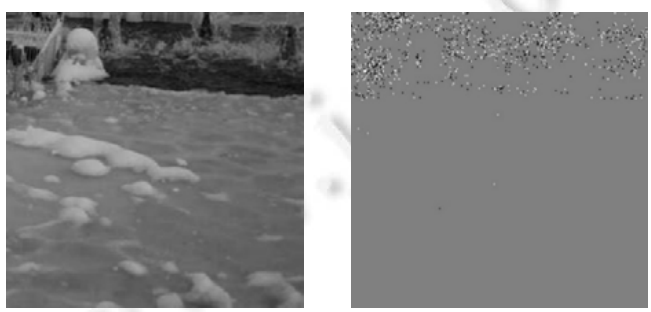
本文实验使用 WOW[5]、S-UNIWARD[6]和 HILL[7]这 3 种常见的空域自适应隐写算法,嵌入率分别为 0.5bpp, 0.4bpp,0.3bpp,0.2bpp,0.1bpp 和 0.05bpp.实验中使用的数据集为 BOSSbase V1.01[27],该数据集包含 10 000 张分辨率为 512×512 像素的灰度图像,图像格式为 pgm.将每张图像分割为 4 张 256×256 像素的灰度图像,这样得到 40 000 张图像.隐写图像集(stego)通过将载体图像集(cover)嵌入秘密信息得到.实验中,将 30 000 张图像用为训练,其中训练集占 80%,验证集占 20%;剩下的 10 000 张图像作为测试集.所有的实验都是在 Windows 10 系统中通过 Caffe1.0[28]深度学习框架实现.

3.2 参数设置

实验中,本文采用随机梯度下降算法(SGD)来训练卷积神经网络.学习率策略(lr_policy)为“step”,stepsize 为 50 000,gamma 为 0.1.基础学习率(base_lr)为 0.001,上一次更新的权重(momentum)为 0.9,权值衰减(weight_decay)为 0.004.由于 GPU 显存的限制,训练阶段每一批次(batch size)设为 16.最大迭代次数(max_iter)为 200 000.所有卷积层的初始化方法为“Xavier”,权重遵循均匀分布,并且保证每层输入和输出的方差保持一致[29].30 个 SRM 滤波器未被标准化.

3.3 实验结果

图 5 展示了 WOW 隐写算法在 0.1bpp 嵌入率下,测试集中一张图像的嵌密结果及其对应嵌入位置.图 5(a)为隐藏信息之后的嵌密图像,图 5(b)为图 5(a)嵌密图像中秘密信息的嵌入位置,其中,白色点表示该位置像素值的进行了+1 修改,黑色点表示该位置像素值进行了–1 修改.模型进行分类后输出的概率值结果分别为 Cover: 0.168572;Stego:0.831528.由于模型将图像判为嵌密图像的概率值更大,因此本文将图 5(a)最终的检测结果判定为嵌密图像.

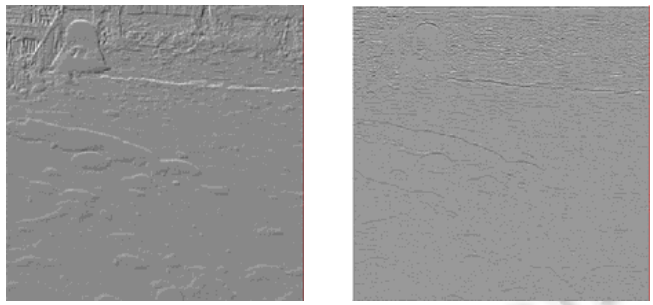


(a) 嵌密图像 (b) 秘密信息嵌入位置

Fig.5 WOW algorithm 0.1bpp embedded effect

图 5 WOW 算法 0.1bpp 嵌密效果图

为了验证模型特征提取的有效性,本文将网络部分层的特征图进行展示.图 6 展示了预处理层进行 SRM 高通滤波核进行滤波处理后的部分噪声残差图.可以看出:不同的滤波核能够从不同的角度提取出携密图像纹理区域和噪声区域的残差信息,并且同时减少了图像内容信息,极大地降低了训练阶段图像内容对隐写噪声特征的学习的影响.图 7 展示了第 1 个卷积层输出的 32 个特征图,这些特征图中的信息依然主要集中在图像的纹理和噪声区域中,说明卷积操作能够从这些秘密信息的主要嵌入区域中有效的提取特征.



(a) “1st”之一滤波结果 (b) “*SQUARE* 5×5”滤波结果

Fig.6 Partial noise residual image outputed from the preprocessing layer

图 6 预处理层输出的部分噪声残差图

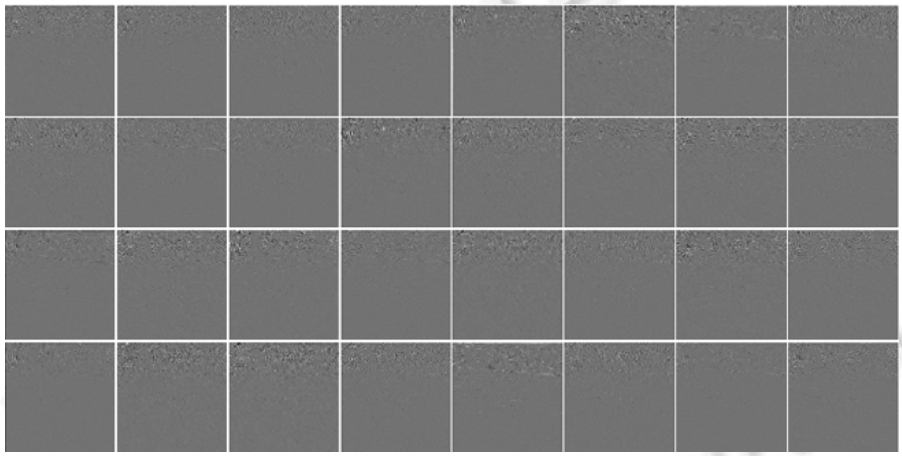


Fig.7 32 feature maps outputted from the first convolutional layer

图 7 第 1 个卷积层输出的 32 个特征图

表 2 中展示了在 WOW 隐写算法下,嵌入率为 0.5bpp,0.4bpp 和 0.3bpp 这 3 种高嵌入率时,现有基于卷积神经网络的隐写分析方法 Pibre-Net[10]、Salomon-Net[11]、Yedroudj-Net[19]和 S-CNN[13],以及本文提出的 Shen-Net 的检测正确率.表 3 展示了 WOW 隐写算法在嵌入率为 0.2bpp、0.1bpp 和 0.05bpp 这 3 种低嵌入率下的检测结果的对比,表中“–”表示该模型在训练阶段未收敛.

Table 2 Comparison of high embedding rate detection accuracy of WOW

表 2 WOW 隐写算法高嵌入率检测准确率对比

Payload(bpp)	0.5	0.4	0.3
Pibre-Net	93.63%	88.43%	81.54%
Salomon-Net	95.63%	93.73%	90.35%
Yedroudj-Net	77.80%	75.18%	69.86%
S-CNN	94.80%	92.41%	88.16%
Shen-Net	97.41%	96.52%	93.44%

Table 3 Comparison of low embedding rate detection accuracy of WOW

表 3 WOW 隐写算法低嵌入率检测准确率对比

Payload(bpp)	0.2	0.1	0.05
Pibre-Net	–	–	–
Salomon-Net	84.57%	73.86%	–
Yedroudj-Net	67.24%	–	–
S-CNN	81.76%	64.53%	51.22%
Shen-Net	89.05%	78.65%	66.55%

从表 2 和表 3 中可以看出,本文提出的 Shen-Net 在 WOW 隐写算法下的检测正确率比现有隐写分析方法更好.与 Pibre-Net 相比,Shen-Net 在嵌入率为 0.5bpp~0.3bpp 时的检测正确率提高了 3%~8%.由于 Pibre-Net 对输入图像直接使用大尺寸卷积核进行卷积操作,使模型无法捕获低嵌入率下的隐写噪声信息,所以在嵌入率为 0.2bpp~0.05bpp 时的训练模型已经无法收敛.与 Salomon-Net 相比,Shen-Net 在嵌入率为 0.5bpp~0.1bpp 时的检测正确率提高了 2%~5%.与 Yedroudj-Net 相比,Shen-Net 在嵌入率为 0.5bpp~0.2bpp 时的检测正确率提高了 20%左右.与 S-CNN 相比,Shen-Net 在嵌入率为 0.5bpp~0.05bpp 时的检测正确率提高了 3%~15%.

特别是在 0.05bpp 下,现有隐写分析方法的网络结构在训练阶段已经难以收敛,S-CNN 网络虽然在训练阶段达到了收敛效果,但是对训练模型进行测试后,检测正确率仅仅只有 51.22%.显然,这个检测效果对于两分类问题很不理想.但是本文提出的 Shen-Net 在 0.05bpp 下,检测正确率能够达到 66.55%.由此可见,本文提出的 Shen-Net 在嵌入率很低的情况下,进行隐写分析也能取得很好的效果.

表 4 和表 5 展示了 S-UNIWARD 隐写算法高嵌入率和低嵌入率时,本文提出的 Shen-Net 与现有隐写分析方法的检测准确率的对比.根据表 4 和表 5 的检测结果可以看出:在 S-UNIWARD 隐写算法下,本文提出的 Shen-Net 的检测性能同样优于现有基于卷积神经网络的隐写分析方法.检测正确率的提升幅度与 WOW 隐写算法检测正确率大致类似.在嵌入率为 0.05bpp 时,Pibre-Net,Salomon-Net 和 Yedroudj-Net 这 3 个网络的训练模型都未能收敛,S-CNN 网络的检测准确率为 58.83%.而本文提出的 Shen-Net 在嵌入率为 0.05bpp 时检测正确率能够达到 73.63%.

Table 4 Comparison of high embedding rate detection accuracy of S-UNIWARD

表 4 S-UNIWARD 隐写算法高嵌入率检测准确率对比

Payload(bpp)	0.5	0.4	0.3
Pibre-Net	94.51%	89.86%	83.06%
Salomon-Net	95.96%	93.10%	91.25%
Yedroudj-Net	75.81%	76.23%	65.70%
S-CNN	95.59%	92.66%	90.34%
Shen-Net	97.86%	96.18%	94.94%

Table 5 Comparison of low embedding rate detection accuracy of S-UNIWARD

表 5 S-UNIWARD 隐写算法低嵌入率检测准确率对比

Payload(bpp)	0.2	0.1	0.05
Pibre-Net	–	–	–
Salomon-Net	86.88%	78.03%	–
Yedroudj-Net	58.30%	–	–
S-CNN	85.88%	74.95%	58.83%
Shen-Net	91.30%	83.18%	73.63%

表 6、表 7 分别展示了 HILL 隐写算法高嵌入率和低嵌入率情况下,Shen-Net 与现有基于卷积神经网络的隐写分析方法的检测准确率的对比.根据表 6 和表 7 的检测结果,本文提出的 Shen-Net 在 HILL 隐写算法下,检测性能同样优于其他 4 个现有基于卷积神经网络的隐写分析方法.在 0.05bpp 下,Pibre-Net,Salomon-Net 和 Yedroudj-Net 这 3 个网络的训练模型依然未能收敛,S-CNN 的测试结果仅为 50.58%,这个检测结果并无太大意义.而本文提出的 Shen-Net 在 HILL 隐写算法 0.05bpp 下,检测准确率还是能够达到 70.48%.

Table 6 Comparison of high embedding rate detection accuracy of HILL

表 6 HILL 隐写算法高嵌入率检测准确率对比

Payload(bpp)	0.5	0.4	0.3
Pibre-Net	94.16%	89.82%	80.19%
Salomon-Net	94.36%	93.22%	90.00%
Yedroudj-Net	77.26%	73.38%	66.72%
S-CNN	94.46%	92.39%	70.61%
Shen-Net	97.10%	96.17%	93.45%

Table 7 Comparison of low embedding rate detection accuracy of HILL

表 7 HILL 隐写算法低嵌入率检测准确率对比

Payload(bpp)	0.2	0.1	0.05
Pibre-Net	–	–	–
Salomon-Net	84.86%	73.73%	–
Yedroudj-Net	61.48%	–	–
S-CNN	82.87%	70.61%	50.58%
Shen-Net	89.52%	80.32%	70.48%

由 WOW,S-UNIWARD 和 HILL 这 3 种常见的隐写算法下的检测性能可见:本文提出的 Shen-Net 相比于现有基于卷积神经网络的隐写分析方法,不仅提升了检测正确率,并且在 0.05bpp 这种其他方法难以检测的低嵌入率下,Shen-Net 同样能够取得较为理想的检测效果.

除了训练模型的检测正确性以外,模型在训练阶段的收敛情况也是评价一个网络的重要指标.loss 值为训练过程中预测结果与实际结果之间的误差,反映了网络在训练阶段的收敛情况.图 8 展示了 Shen-Net 与其他 4 个网络在训练阶段 loss 值的变化情况,训练集的隐写算法为 WOW,嵌入率为 0.3bpp.根据 loss 曲线能够看出:相比其他 4 个网络,Shen-Net 在训练阶段能够明显地快速进行收敛.在 30 000 次左右的迭代时,loss 值已基本保持在极低范围内,而其他网络的 loss 曲线基本都还在收敛阶段,且 loss 值都远远高于 Shen-Net 的 loss 值.

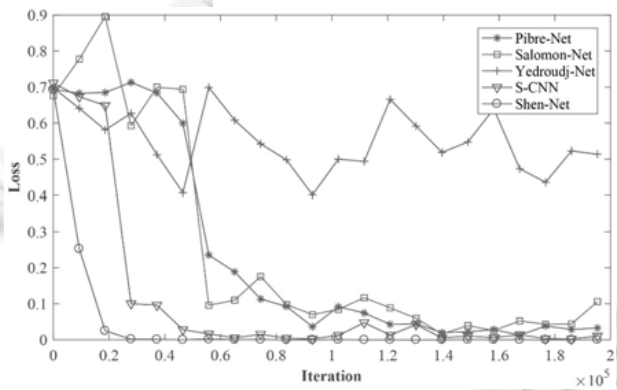


Fig.8 Loss variation of 0.3bpp embedding rate training stage of WOW

图 8 WOW 隐写算法 0.3bpp 训练阶段 loss 变化情况

在卷积神经网络中,模型的训练与测试所耗时间是衡量模型性能重要指标,表 8 展示了本章提出的 Shen-Net 与对比的 4 个网络在训练与测试阶段所耗费的时间.其中,每个网络在训练阶段的迭代次数都为 20 万次,测试结果为对一张图像进行检测所耗费时间.

Table 8 Performance during training and testing

表 8 训练与测试的性能比较

	训练(迭代 20 万次/h)	测试(一张图像/s)
Pibre-Net	10.5h	1s
Salomon-Net	1.3h	0.1s
Yedroudj-Net	14.5h	2.5s
S-CNN	7.6h	0.7s
Shen-Net	4h	0.4s

Yedroudj-Net 由于其网络深度相对较大,模型的参数较多,因此训练阶段需要耗费 14.5 小时.Salomon-Net 的网络深度为两层卷积结构,网络结构较为简单,且第 1 次卷积层只有一个卷积核,从而模型参数较少,因此训练阶段只需 1.3 小时即可完成.Shen-Net 与其他 4 个网络相比,在加深网络层数的基础上,通过对卷积核的设计对模型参数数量进行了一定限制,因此训练阶段所耗费时间较为适中.测试阶段所耗时间与训练时间相对应,Yedroudj-Net 运行时间同样为最长,Salomon-Net 只需 0.1 秒即可完成,而 Shen-Net 对单张图片的测试时间为 0.4

秒.根据 Shen-Net 与其他 4 个网络的训练和测试时间可以反映出,S-CNN 的模型性能接近于 5 个模型的平均值,Shen-Net 的模型性能虽然不是最优的,但是也能达到较好的水平.

3.4 迁移学习实验

为了进一步对提高低嵌入率的检测效果,如 0.05bpp、0.1bpp 和 0.2bpp,本文采用迁移学习方法将嵌入率为 0.3bpp、0.4bpp 和 0.05bpp 训练得到的网络模型参数,分别迁移至 0.05bpp、0.1bpp 和 0.2bpp 下进行微调训练.通过将高嵌入率的模型参数有效地迁移至相同隐写算法低嵌入率中进一步进行特征学习,有效提升了模型对低嵌入率的隐写分析能力.除了通过直接将嵌入率为 0.3bpp、0.4bpp 和 0.05bpp 下的预训练的模型迁移至低嵌入率中进行训练以外,实验中通过逐步迁移(step by step)学习,对嵌入率差距较大的情况进行有效的参数迁移.针对本文提出的 Shen-Net 网络,分别对 WOW、S-UNIWARD 和 HILL 这 3 种隐写算法进行迁移学习实验.

表 9 展示了 Shen-Net 对 WOW 隐写算法未进行迁移学习与 4 种迁移学习方式的检测准确率对比.对 0.05bpp 进行迁移学习,较好地提升了准确率,其中,通过 0.3bpp 和逐步迁移的方式,相比未进行迁移学习提升了 2%左右.其次,逐步迁移的方法对 0.1bpp 的检测准确率也提升了 1%左右.但是对于 0.2bpp 而言,迁移学习并没有提升检测准确率.

Table 9 Comparison of transfer learning detection accuracy of WOW

表 9 WOW 隐写算法迁移学习检测准确率对比

Payload(bpp)	0.2	0.1	0.05
No-transfer	89.05%	78.65%	66.55%
Trans-0.5bpp	88.94%	78.85%	67.45%
Trans-0.4bpp	89.04%	78.69%	67.21%
Trans-0.3bpp	89.04%	78.78%	68.62%
Trans-step-by-step	89.05%	79.61%	68.45%

S-UNIWARD 隐写算法下,使用迁移学习的检测准确率对比结果见表 10.在 0.05bpp 下,通过迁移学习能够在未进行迁移学习的基础上提升 1%左右,其中,0.5bpp 和 0.4bpp 迁移学习的效果最好.在 0.1bpp 和 0.2bpp 下,迁移学习的方法检测准确率都能得到一定的提升.

Table 10 Comparison of transfer learning detection accuracy of S-UNIWARD

表 10 S-UNIWARD 隐写算法迁移学习检测准确率对比

Payload(bpp)	0.2	0.1	0.05
No-transfer	91.30%	83.18%	73.63%
Trans-0.5bpp	91.66%	83.65%	74.68%
Trans-0.4bpp	91.73%	83.99%	74.68%
Trans-0.3bpp	91.66%	83.65%	74.52%
Trans-step-by-step	91.53%	83.87%	74.62%

迁移学习方法对 HILL 隐写算法检测准确率的对比结果见表 11,整体而言,逐步迁移学习的检测效果能够得到最好的提升.其中,在 0.05bpp 和 0.1bpp 下,逐步迁移的方法能够在未进行迁移学习的基础上分别提升 2%和 1%左右.

Table 11 Comparison of transfer learning detection accuracy of HILL

表 11 HILL 隐写算法迁移学习检测准确率对比

Payload(bpp)	0.2	0.1	0.05
No-transfer	89.52%	80.32%	70.48%
Trans-0.5bpp	89.73%	81.18%	70.99%
Trans-0.4bpp	89.65%	80.39%	70.28%
Trans-0.3bpp	89.56%	81.39%	71.62%
Trans-step-by-step	89.75%	81.41%	72.10%

通过以上实验可以发现,使用逐步迁移学习的方法能够获得更为稳定的准确率提升.特别是在 0.05bpp 下,相比未进行迁移学习的效果提升更为明显.但是在 0.2bpp 下,由于 Shen-Net 未进行迁移学习时,在训练阶段同样能够学习到足够的隐写特征,因此迁移学习对检测准确率的提升不大.

4 结束语

本文针对现有空域隐写分析方法在低嵌入率下难以区分的问题,通过分析现有基于卷积神经网络的隐写分析方法的特点,构造了一个新的网络结构 Shen-Net.实验结果证明:新提出的网络结构在对 WOW,S-UNIWARD 和 HILL 这 3 种常见空域内容自适应隐写算法进行隐写分析时,准确率得到了较高的提升.在嵌入率较低的情况下,现有网络结构无法收敛或准确性很低,而本文设计的网络结构仍能够取得较为理想的检测准确率.此外,本文还通过采用逐步迁移学习的方法进一步提高了对低嵌入容量的检测准确率.由于在现实生活中,JPEG 格式图像的使用更为常见,下一步我们将对 JPEG 格式图像的隐写分析方法进行深入研究.

References:

[1] Huang W, Zhao XF, Feng DG, Sheng RN. JPEG steganalysis based on feature fusion by principal component analysis. Ruan Jian Xue Bao/Journal of Software, 2012,23(7):1869−1879 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4107.htm [10.3724/SP.J.1001.2012.04107]

[2] Zhang YW, Zhang WM, Yu NH. Specific testing sample steganalysis. Ruan Jian Xue Bao/Journal of Software, 2018,29(4): 987−1001 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5411.htm [doi: 10.13328/j.cnki.jos.005411]

[3] Huang DZ, Zhang JF, Zhang R, Li PC, Guo YB. New system of multi-modal information hiding based on big data environment. Journal of Electronic, 2017,45(2):477−484 (in Chinese with English abstract).

[4] Tang GM, Sun Y, Xu XY, Wang Y. Adaptive JPEG steganography based on distortion cost updating. Journal on Communications, 2017,38(9):1−8 (in Chinese with English abstract).

[5] Holub V, Fridrich J. Designing steganographic distortion using directional filters. In: Proc. of the IEEE Int'l Workshop on Information Forensics and Security. IEEE, 2012. 234−239.

[6] Holub V, Fridrich J, Denemark T. Universal distortion function for steganography in an arbitrary domain. EURASIP Journal on Information Security, 2014,2014(1):1.

[7] Li B, Wang M, Huang J, Li X. A new cost function for spatial image steganography. In: Proc. of the IEEE Int'l Conf. on Image Processing. Paris: IEEE, 2014. 4206−4210.

[8] Fridrich J, Kodovsky J. Rich models for steganalysis of digital images. IEEE Trans. on Information Forensics and Security, 2012, 7(3):868−882.

[9] Zhai LM, Jia J, Ren WX, Xu YB, Wang LN. Recent advances in deep learning for image steganography and steganalysis. Journal of Information Security, 2018,3(6):2−12 (in Chinese with English abstract).

[10] Pibre L, Pasquet J, Ienco D. Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover sourcemismatch. Electronic Imaging, 2016,2016(8):1−11.

[11] Salomon M, Couturier R, Guyeux C, Coucho JF, Bahi JM. Steganalysis via a convolutional neural network using large convolution filters for embedding process with same stego key: A deep learning approach for telemedicine. European Research in Telemedicine/La Recherche Européenne en Télémédecine, 2017,6(2):79−92.

[12] Pevný T, Filler T, Bas P. Using high-dimensional image models to perform highly undetectable steganography. In: Proc. of the Int'l Workshop on Information Hiding. Berlin: Springer-Verlag, 2010. 161−177.

[13] Gao PX, Wei LX, Liu J, Liu MM. Image steganalysis based on convolution neural network. Application Research of Computers, 2019,36(1):54−288 (in Chinese with English abstract).

[14] Xu G, Wu HZ, Shi YQ. Structural design of convolutional neural networks for steganalysis. IEEE Signal Processing Letters, 2016, 23(5):708−712.

[15] Tan S, Li B. Stacked convolutional auto-encoders for steganalysis of digital images. In: Proc. of the Signal and Information Processing Association Annual Summit and Conf. Chiang Mai: Asia-Pacific, 2014. 1−4.

[16] Qian Y, Dong J, Wang W, Tan T. Deep learning for steganalysis via convolutional neural networks. In: Alattar AM, *et al*, eds. Proc. of SPIE-IS&T, 2015. 9409: 94090J.

[17] Qian Y, Dong J, Wang W, Tan T. Learning and transferring representations for image steganalysis using convolutional neural network. In: Proc. of the IEEE Int'l Conf. on Image Processing. 2016. 2752−2756.

[18] Xu G, Wu HZ, Shi YQ. Ensemble of CNNs for steganalysis: An empirical study. In: Proc. of the ACM Workshop on Information Hiding and Multimedia Security. ACM, 2016. 103−107.

[19] Yedroudj M, Comby F, Chaumont M. Yedroudj-net: An efficient CNN for spatial steganalysis. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. IEEE, 2018. 2092−2096.

[20] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. of the Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770−778.

[21] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proc. of the Int'l Conf. on Machine Learning. IMLS, 2015. 448−456.

[22] Ye J, Ni J, Yi Y. Deep learning hierarchical representations for image steganalysis. IEEE Trans. on Information Forensics and Security, 2017,12(11):2545−2557.

[23] Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. Multimedia tools and applications, 2018,77(9):10437−10453.

[24] Tsang CF, Fridrich J. Steganalyzing images of arbitrary size with CNNs. Electronic Imaging, 2018,2018(7):1−8.

[25] Shen J. Research on image steganalysis with low embedding rate based on convolutional neural network [MS. Thesis]. Changsha: Hunan University, 2020 (in Chinese with English abstract).

[26] Chen M, Sedighi V, Boroumand M, Fridrich J. JPEG-phase-aware convolutional neural network for steganalysis of JPEG images. In: Proc. of the ACM Workshop on Information Hiding and Multimedia Security. Denver: ACM, 2017. 75−84.

[27] Bas P, Filler T, Pevný T. "Break our steganographic system": The ins and outs of organizing BOSS. In: Proc. of the Int'l Workshop on Information Hiding. Berlin: Springer-Verlag, 2011. 59−70.

[28] Jia Y, Shelhamer E, Donahue J, Karayev S. Caffe: Convolutional architecture for fast feature embedding. In: Proc. of the ACM Int'l Conf. on Multimedia. Glasgow: ACM, 2014. 675−678.

[29] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proc. of the Int'l Conf. on Artificial Intelligence and Statistics. 2010. 249−256.

附中文参考文献:

[1] 黄炜,赵险峰,冯登国,盛任农.基于主成分分析进行特征融合的 JPEG 隐写分析.软件学报,2012,23(7):1869−1879. http://www.jos.org.cn/1000-9825/4107.htm [10.3724/SP.J.1001.2012.04107]

[2] 张逸为,张卫明,俞能海.针对特定测试样本的隐写分析方法.软件学报,2018,29(4):987−1001. http://www.jos.org.cn/1000-9825/5411.htm [doi: 10.13328/j.cnki.jos.005411]

[3] 黄殿中,张静飞,张茹,李鹏超,郭云彪.基于大数据环境的多模态信息隐藏新体系.电子学报,2017,45(2):477−484.

[4] 汤光明,孙艺,徐潇雨,王宇.动态更新失真代价的自适应 JPEG 隐写算法.通信学报,2017,38(9):1−8.

[9] 翟黎明,嘉炬,任魏翔,徐一波,王丽娜.深度学习在图像隐写术与隐写分析领域中的研究进展.信息安全学报,2018,3(6):2−12.

[13] 高培贤,魏立线,刘佳,刘明明.基于卷积神经网络的图像隐写分析方法.计算机应用研究,2019,36(1):54−288.

[25] 沈军.基于卷积神经网络的低嵌入率图像隐写分析研究[硕士学位论文].长沙:湖南大学,2020.

沈军(1993－),男,硕士,主要研究领域为多媒体信息安全.

秦拯(1969－),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为数据安全与隐私保护,云计算,大数据,机器学习.

廖鑫(1985－),男,博士,副教授,博士生导师,CCF 高级会员,主要研究领域为多媒体安全,人工智能安全,数字取证,密码学.

刘绪崇(1974－),男,博士,教授,主要研究领域为信息网络安全,网络犯罪侦查,人工智能.