Deep Learning Hierarchical Representations for Image Steganalysis【Ye-Net：图像隐写分析的深度学习层次表示】

原创

CV误会了我 于 2021-11-01 21:27:42 发布 637 收藏 8

文章标签： 深度学习 r语言 计算机视觉

本文链接：https://blog.csdn.net/wangsanNOLOVE/article/details/121083331

Deep Learning Hierarchical Representations for Image Steganalysis

【Ye-Net：图像隐写分析的深度学习层次表示】

Abstract

目前流行的数字图像隐写通信检测方法主要包括残差计算、特征提取和二值分类三个步骤。在本文中，我们提出了一种基于卷积神经网络(CNN)的数字图像隐写分析替代方法，该方法能够在一个统一的框架中很好地复制和优化这些关键步骤，并直接从原始图像中学习层次表示。本文提出的CNN与传统计算机视觉任务中使用的CNN结构有很大的不同。**本文所提出的CNN第一层的权值不是采用随机策略，而是使用在空间丰富模型(SRM)中残差映射计算中使用的基本高通滤波集初始化，作为正则化器有效地抑制图像内容。为了更好地捕捉嵌入信号的结构，我们在CNN模型中采用了一种叫做截断线性单元的新激活函数**。最后，我们进一步提高了基于cnn的隐写分析器的性能。采用WOW、S-UNIWARD和HILL三种最先进的空间隐写算法对模型的有效性进行了评估。与SRM和它的选择通道感知变体maxSRMd2相比，我们的模型在各种有效载荷下的所有测试算法中实现了优越的性能。

I. INTRODUCTION

隐写术是通过稍微修改像素值(空间域)或DCT系数(JPEG域)来隐藏图像中秘密信息的科学和艺术。目前，最安全的隐写方案是内容自适应隐写方案，将秘密数据保存在内容复杂的区域，且嵌入痕迹不易被检测到。空间领域的例子包括HUGO [1]，WOW[2]和S-UNIW ARD[3]。

与图像隐写学的发展相对应，旨在揭示图像中隐藏信息存在的隐写分析也取得了长足的进展。目前最先进的空间隐写分析器是空间丰富模型(SRM)[4]及其变体[5]、[6]。这些隐写分析工具是通过将一个丰富的模型组合成多个不同的子模型的联合，这些子模型是由使用线性和非线性高通滤波器获得的量化图像噪声残差的相邻样本联合分布形成的。近年来，为了更好地应对日益增长的安全性和内容自适应隐写方案，提出了一些感知选择信道的隐写分析特征集[6]、[7]，其中，基于富媒体模型，利用选择信道，maxSRM[6]在不同程度上提高了空间域所有内容自适应隐写方案的检测能力。

目前，最好的图像隐写分析器都是使用基于特征的隐写分析和机器学习来构建的，并且共享相同的流程:即噪声残差计算、特征构建和二值分类。基于特征的隐写分析的成功在很大程度上依赖于特征工程的过程，即使用领域知识，如掩护源模型及其对手的行为，来创建使机器学习算法工作的特征。对于上述的管道，残差有助于提高隐写信号的信噪比，最先进的特征集是不同滤波器残差的共现并集，即所谓的富模型，往往是高维的(例如，30,000或更多)。从隐写分析的角度来看，为了获得更完整的cover源描述，高维表示是必然的趋势，这表明用于隐写分析的特征越来越复杂。此外，请注意，目前最先进的隐写分析特征是启发式设计的，分类器的优化与特征提取步骤无关。换句话说，隐写分析的管道在一个统一的框架中几乎没有得到优化。

在本文中，我们证明了隐写分析的过程可以通过深度卷积神经网络(CNN)[8]来交替实现，以学习图像隐写分析的优化深度层次表示。CNN的一个重要属性是它可以从高维感觉输入和有效地提取复杂的统计依赖关系学习的深度(分层)表示,重用和中间相结合的概念,允许它概括在一个各种各样的计算机视觉(CV)的任务,包括图像分类[9],人脸识别[10]等。这自然促使我们考虑训练一个CNN来区分cover和stego。这样，待检测的原始图像可以使用训练过的CNN直接映射到二进制标签(cover或stego)。此外，特征提取可以与分类器一起优化，这有助于我们从复杂的特征设计步骤中解脱出来。

Qian的工作[11]首次尝试使用CNN进行隐写分析，其中提出了一种用于空间域图像隐写分析的高斯神经元卷积神经网络（GNCNN）。通过使用高斯函数代替传统CNN中的ReLU或sigmoid作为激活函数，GNCNN实现了与BOSSbase上的SRM相当的性能[12]。最近，Xu[13]研究了用于图像隐写分析应用的CNN结构设计，其特点是（1）在第一卷积层中嵌入绝对激活（ABS）层，以改进后续层中的统计建模；（2） 在网络的早期阶段应用TanH激活函数以防止过度拟合；（3） 在每个非线性激活层之前立即执行批量归一化（BN）。他们的结果表明，设计良好的CNN有可能在隐写分析中提供更好的检测性能。作者在[14]晚些时候扩展了他们的工作，将[13]中的网络用作集成分类器的基础学习器，并获得了可与SRM相媲美的结果。

在本文中，我们开发了一个特定于隐写分析应用的有监督的CNN模型。所提出的CNN与其他CNN有几个突出的不同之处，总结如下:(1)所提出的CNN的第一层作为噪声残差计算的预处理模块。用SRM[4]中剩余映射的30个基本滤波器初始化第一层的权值，而不是采用随机策略，这相当于第一层的30个输出特征映射，有助于加速网络的收敛。(2)在本文所提出的CNN中，我们采用了一组混合激活函数，其中，除了传统的ReLU函数外，在网络的前几层引入了一个叫做截断线性单元(TLU)的新函数。实际上，与传统的CV任务不同的是，隐写过程可以看作是在覆盖层上添加极低信噪比的嵌入信号的过程。前几层采用TLU有助于适应嵌入信号的分布，并使CNN更有效地学习高通滤波器。(3)最后，我们利用所提出的CNN训练中的选择通道，进一步提高了隐写分析性能。所提出的CNN的有效性通过使用几种最先进的隐写工具对各种有效载荷进行彻底实验的证据得到了验证。与之前基于CNN的隐写分析器[11]相比，本文提出的CNN在检测精度方面取得了相当大的性能提升，并明显优于当前最先进的手工特性集，如SRM[4]和maxSRMd2[6]。

本文的其余部分组织如下。在第二节中，我们简要回顾了目前流行的空间域图像隐写分析方法和卷积神经网络(CNNs)的框架。第三节描述了本文所提出的CNN的结构，第四节给出了实验结果和分析，第五节给出了结论。

II. PRELIMINARIES

A. The Framework of Prevailing Image Steganalysis Methods

图像隐写分析的成熟范式[4]，[15]，[16]主要包括噪声残差计算、特征提取和二值分类三个主要步骤，如图1所示。

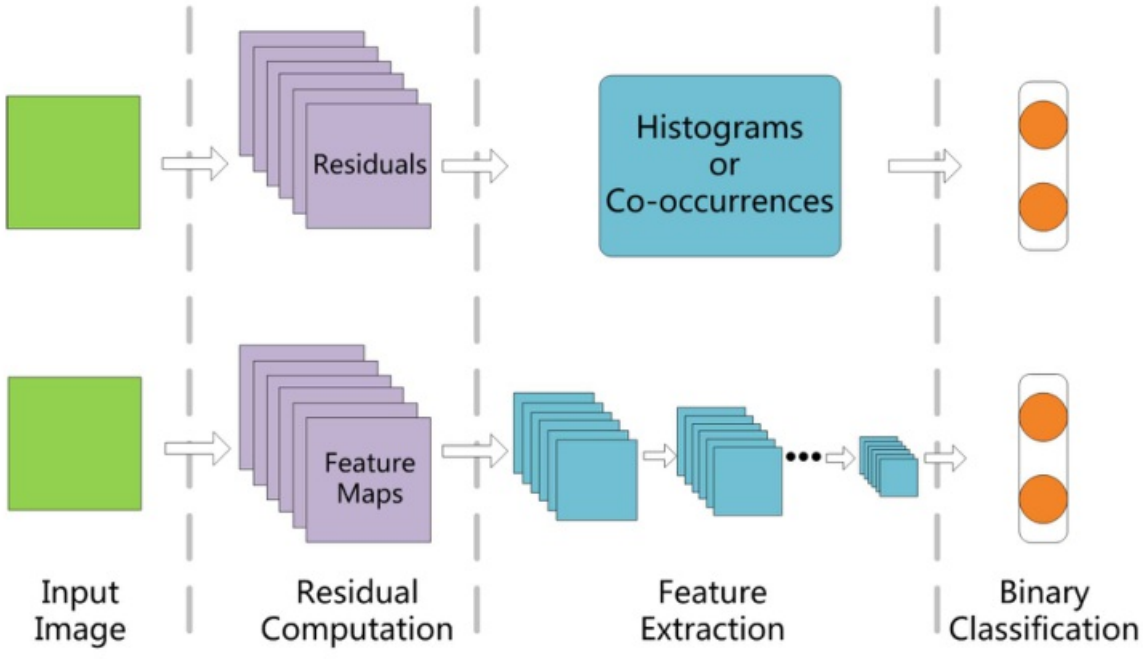


Fig. 1. The framework of image steganalysis methods and its similarity with the convolutional neural networks.

1、噪声残差计算:隐写术中的嵌入操作可以看作是在覆盖层上添加极低振幅的噪声。因此，在隐写分析中，采用噪声残差模型代替原始像素模型更为明智。这一思想最初在[17]中提出，后来在[4]、[15]、[16]、[18]等方法中被采用和发展。对于测试图像X=(xij)，隐写分析中一种流行的策略是从像素预测器中计算噪声残差R=(rij):

1) Noise Residual Computation: The embedding operation in steganography can be viewed as adding extremely low amplitude noise to the cover. Therefore, it is wiser to model the noise residuals instead of raw pixels in steganalysis. Such an idea was initially proposed in [17] and was later adopted and developed in several subsequent methods [4], [15], [16], [18]. For a test image $X = (x_{ij})$, a popular strategy in steganalysis is to compute the noise residuals $R = (r_{ij})$ from a pixel predictor:

$$r_{ij} = Pred(N(x_{ij})) - lx_{ij}, \tag{1}$$

where $N(x_{ij})$ is a set of neighboring pixels of x_{ij}, $l \in \mathbb{N}$ is the residual order, and $Pred(\cdot)$ is the adopted predictor. In practice, many steganalysis schemes [17], [19] implement the predictor by convolving a finite impulse response filter K' with image X:

$$R = X * K' - lX = (r_{ij}) = (\sum_{r,c} x_{ij}^{rc} k'^{rc} - lx_{ij}), \tag{2}$$

where $*$ denotes the convolution operator, and r, c are the index of the kernel K'. According to the distributive law, the residuals above can be reformulated as:

$$R = X * K = (r_{ij}) = (\sum_{r,c} x_{i,j}^{r,c} k^{rc}). \tag{3}$$

在隐写分析中，滤波器有很多选择(线性或非线性)，可以用来生成不同的残差，并捕获相邻像素之间的不同依赖关系。残差的多样性是所谓富媒体模型(RM)成功的基础。

2、特征提取:这在隐写分析中至关重要。有了更多的鉴别特征，就可以更容易地区分掩蔽图像和隐写图像。在这一步中，通过直方图或共现来模拟相邻残差的联合或条件概率分布。对于SRM及其变体，特征是建立在四阶共现矩阵的基础上。以水平共生为例，我们有:

$$c^h_{d_0 d_1 d_2 d_3} = \sum_{i,j=1}^{n_1, n_2 - 3} [r_{i,j+k} = d_k, \quad \forall k = 0, 1, 2, 3] \cdot \varphi(\beta_{i,j})$$
$$d_k \in \{-Tq, (-T+1)q, \cdots, Tq\}, \tag{4}$$

其中[·]为艾弗森括号，当语句inside为true时，其结果为1，否则为0。其中ϕ(βi,j)是相应嵌入概率的统计度量。对于不同版本的SRM，其中的ϕ(βi,j)有不同的值:

$$\varphi(\beta_{i,j}) = \begin{cases} 1, & \text{SRM} \\ max(2\beta_{i,j+k}), k = 0, 1, 2, 3, & \text{maxSRM} \\ [\beta_{ij} \geq \beta_{threshold}], & \text{tSRM}. \end{cases} \quad (5)$$

注意，maxSRM和tSRM中的ϕ(βi,j)s从一个地方到另一个地方都不同，这表明它们都是选择信道感知的。

2) Feature Extraction: This is critical in steganalysis. With more discriminative features, it would be much easier to distinguish cover images from stego ones. In this step, the joint or conditional probability distributions of neighboring residuals are modeled through histograms or co-occurrences. For SRM and its several variants, the features are built on the basis of fourth order co-occurrence matrixes. Take horizontal co-occurrence for example, we have:

$$c^h_{d_0 d_1 d_2 d_3} = \sum_{i,j=1}^{n_1, n_2 - 3} [r_{i,j+k} = d_k, \quad \forall k = 0, 1, 2, 3] \cdot \varphi(\beta_{i,j})$$
$$d_k \in \{-Tq, (-T+1)q, \cdots, Tq\}, \quad (4)$$

where [·] is Iverson bracket whose result is 1 when statement inside is true and 0 otherwise. And $\varphi(\beta_{i,j})$ is a statistical measure of the corresponding embedding probability. For different versions of SRM, there are different values for $\varphi(\beta_{i,j})$:

$$\varphi(\beta_{i,j}) = \begin{cases} 1, & \text{SRM} \\ max(2\beta_{i,j+k}), k = 0, 1, 2, 3, & \text{maxSRM} \\ [\beta_{ij} \geq \beta_{threshold}], & \text{tSRM}. \end{cases} \quad (5)$$

Note that the $\varphi(\beta_{i,j})$s in maxSRM and tSRM vary from one place to another, indicating that both of them are selection-channel-aware.

3、二值分类:隐写分析的最后一步是使用精心设计的分类器(支持向量机(SVM)或集成分类器)将图像分类为掩体或隐写体，在实际应用前需要通过监督学习进行训练

B. Convolutional Neural Network Architecture

卷积神经网络由一个或几个卷积层组成，随后是一些完全连接的神经元层。卷积层的输入和输出是一组称为feature map的数组，而每个卷积层通常通过卷积、非线性激活和池化三步来生成feature map。第一步是使用已知特征映射的内核进行一些过滤。因此，每一个kernel都应用于前一层生成的现有feature map上。设用Fn(X)表示n层的输出特征映射，其核(滤波)和偏置分别由Wn和Bn定义，有:

$$F^n(X) = pooling(f^n(F^{n-1}(X) * W^n + B^n)), \quad (6)$$

where $F^0(X) = X$ is the input data, $f^n(\cdot)$ is a non-linear activation function that applies to each element of its input, e.g., TanH or ReLU function, and $pooling(\cdot)$ represents the pooling operation, including mean-pooling or max-pooling, etc. Generally speaking, the non-linear activation and pooling operation are optional in a specific layer. For a classification problem, a complete network usually contains several cascaded convolutional layers and ends with one fully-connected layer followed by a softmax classifier.

其中F0(X)=X为输入数据，fn(·)为非线性激活函数，应用于其输入的每个元素，如TanH或ReLU函数，pooling(·)表示pooling操作，包括mean-pooling或max-pooling等。一般来说，非线性激活和池操作在特定层是可选的。对于一个分类问题，一个完整的网络通常包含几个级联的卷积层，并以一个全连接层和一个softmax分类器结束。

显然，上述CNN模型可以很好地模拟现代隐写分析框架中的三个关键步骤。根据(3)，残差计算实际上是通过卷积来实现的，卷积可以通过一个卷积层来实现。CNN中的多重卷积层的级联可以被训练来学习或提取高层和歧视表示或原始数据的特征,这就解释了CNN的成功在许多图像和视频识别的问题,这也同时在隐写式密码解密的目标特征提取。在分类步骤上，CNN中的softmax分类器类似于SVM或集成分类器。事实上，基于CNN的隐写分析器可以在一个独特的架构中自动统一残差计算、特征提取和分类步骤，而无需事先进行任何特征选择，并作为一个整体框架同时进行优化。

III.基于卷积神经网络的隐写分析

从第二节的分析可以看出，CNN模型能够很好地模拟现代隐写分析框架中的三个关键步骤。因此，利用卷积神经网络开发图像隐写分析器不仅是可行的，而且是可行的。然而，隐写分析任务与计算机视觉任务有很大的不同，在计算机视觉中，cnn已经取得了巨大的成功。隐写分析中需要处理的隐写噪声通常是人类感知系统无法感知的。事实上，在精心设计的隐写方案中，隐写通常不仅在视觉上，而且在统计上都与掩蔽物非常相似。因此，基于CNN的隐写分析器的特征表示应该与传统的CV任务有很大的不同。因此，当将CNN训练为隐写分析器时，具有随机初始权值的CNN通常不能收敛也就不足为奇了(见表I)。

TABLE I

THE PERFORMANCE OF DIFFERENT INITIALIZATION STRATEGIES OF THE FIRST CONVOLUTIONAL LAYER IN TERMS OF DETECTION ERROR (P_E) FOR RELU-CNN AND TLU-CNN($T = 3$) ON THREE STEGANOGRAPIC SCHEMES AT A PAYLOAD OF 0.2 BPP ON RESAMPLED IMAGES. THE INVOLVED NETWORKS ARE TRAINED ON BOSS+BOWS2+AUG AND TESTED ON BOSS_TEST

Algorithm	Model	Random	Fixed	Learned
WOW	ReLU-CNN	0.5	0.2259	0.2136
	TLU-CNN	0.5	0.2261	**0.1982**
S-UNIWARD	ReLU-CNN	0.5	0.2968	0.2937
	TLU-CNN	0.5	0.2807	**0.2540**
HILL	ReLU-CNN	0.5	0.2980	0.2971
	TLU-CNN	0.5	0.3068	**0.2761**

因此，为了将领域知识融入到基于CNN的隐写分析器的学习中，需要一些特定于隐写分析的定制设计。

A. The Architecture

如图2所示，所提出的CNN由10层组成，并以具有双向softmax的完全连接层结束，其产生超过2个类别标签的分布。在每次卷积运算后应用非线性激活。并且从第一层到第三层的池操作被抑制。与使用两个或多个完全连接层的其他传统CNN架构不同，我们在网络末端仅使用一个必要的双向完全连接层。这是因为完全连接的层通常涉及太多要训练的参数，这很容易导致过度拟合，特别是当训练集不够大时，我们的任务就是这样。此外，除了图2中所示的层之外，没有在网络中使用的其他层，例如本地响应规范化（LRN）[9]、丢失[9]、批量规范化（BN）[20]或本地对比度规范化（LCN）[21]。虚线框内的组件显示了探索选择通道知识时的前两个操作，将在第Ⅲ-D节中详细介绍。网络的深度和宽度以及滤波器的大小通过基于性能和模型复杂性之间权衡的实验确定。

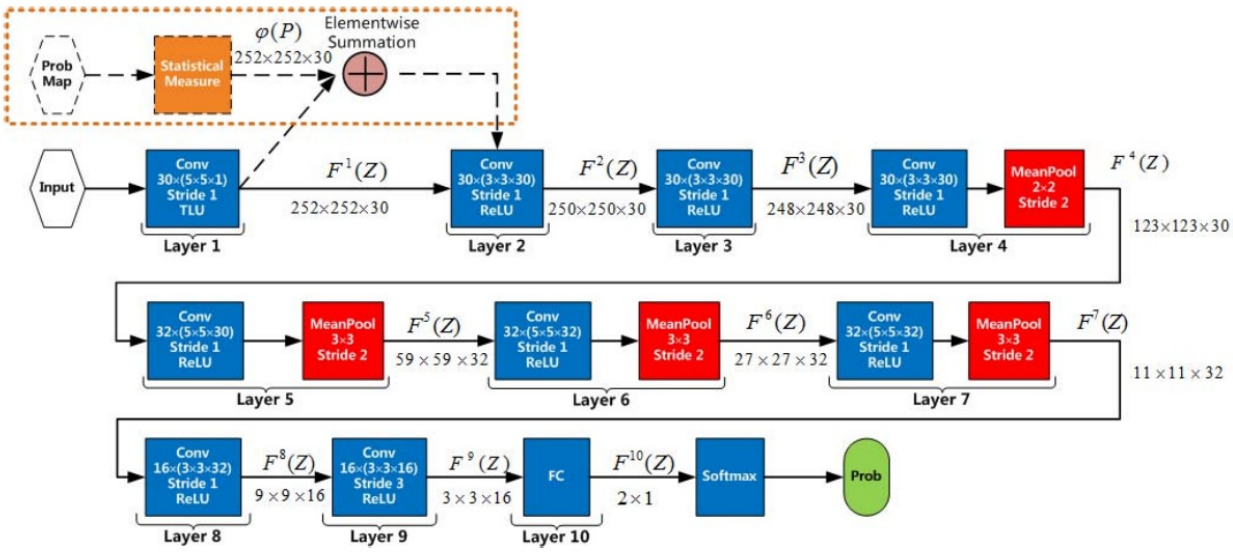


Fig. 2. The architecture of the proposed 10-layer convolutional neural network. For each convolutional layer, the input feature maps are the output of its previous layer. The layers in the dotted box only exist in the selection-channel-aware version of the proposed network.

然后，我们继续讨论CNN的命名约定，以及我们在所提出的模型上重复实验的方式，这将贯穿本文。对于非线性激活函数为ReLU的网络，我们将其称为ReLU CNN，而如果一些激活函数被新引入的截断线性单元（TLU）所取代，这将在第Ⅲ-B节中详细阐述，则称为TLU-CNN。此外，当嵌入概率的统计度量被纳入网络设计时，TLU-CNN成为其选择信道感知版本SCA-TLU-CNN。对于CNN模型上的所有实验，我们使用三种不同的训练/验证/测试集重复训练和测试程序三次。最后的实验结果是通过平均三个测试结果得到的。除非另有规定，本文中涉及的实验网络使用BOSS + BOWS2 + AUG数据集上的重采样图像进行训练，并使用BOSS_测试上的重采样图像进行测试（详情见第Ⅳ-B节）。

在本节的其余部分中，我们将详细阐述在设计我们提出的卷积神经网络时所采用的几个关键技术

B. Initialization With High-Pass Filters in SRM

在SRM中使用高通滤波器初始化

如前所述，残差的计算可以通过卷积层很好地模拟。受此启发，我们使用SRM中使用的高通滤波器核而不是随机值初始化卷积神经网络中第一卷积层的权重。尽管在Qian等人的工作[11]中也使用了这种策略，但在GNCNN模型中，只有SRM中的“平方5×5”滤波器用于初始化其第一层。根据我们的理解，SRM中的残差有助于提高信噪比 (stego signal to image content)（隐藏信号到图像内容），正是不同滤波器残差模型的组合使得rich模型（RM）在隐写分析中取得了成功。因此，我们建议增加第一卷积层的宽度，并使用SRM中用于计算残差映射的所有30个基本线性滤波器（“SPAM”滤波器及其旋转对应物）的内核初始化权重。

上述基本滤波器对应于SRM中的7个剩余类，包括“1”类中的8个滤波器、“2”类中的4个滤波器、“3”类中的8个滤波器、“3×3”类中的1个滤波器、“5×5”类中的1个滤波器、“3×3”类中的4个滤波器和“5×5”类中的4个滤波器，总共30个基本滤波器的最大内核大小为5×5。因此，我们将CNN all的第一个卷积层中加权矩阵的核大小设置为5×5，如图2所示。假设W5×5 CNN和Wm×n 分别作为SRM中的权值矩阵和滤波核，我们用WSRM初始化WCNN的中心部分，将WCNN的剩余元素保留为零。换句话说，我们将wm×n SRM设为带零的w5×5cnn。值得注意的是，对于所有SRM过滤器，我们不会通过除以公式（1）中的剩余顺序来对其进行规范化。

The basic filters above correspond to 7 residual classes in SRM, which include 8 filters in class “1st,” 4 in class “2nd,” 8 in class “3rd,” 1 in class “SQUARE 3×3,” 1 in class “SQUARE 5×5,” 4 in class “EDGE 3×3” and 4 in class “EDGE 5×5,” for a total of 30 basic filters with maximum kernel size of 5×5. Therefore, we set the kernel size of weighting matrix in the first convolutional layer of our CNN all to 5×5 as shown in Fig. 2. Let $W_{CNN}^{5\times5}$ and $W_{SRM}^{m\times n}$ be the weight matrix and filter kernel in SRM, respectively, we initialize the central part of W_{CNN} with W_{SRM} and leave the remaining elements of W_{CNN} to be zeros. In another word, we pad $W_{SRM}^{m\times n}$ to be the $W_{CNN}^{5\times5}$ with zeros. It is worth noting that for all the SRM filters, we do not normalize them by dividing the residual orders l in formula (1).

上述初始化策略作为机器学习中的正则化项，极大地缩小了可行参数空间，有助于网络的收敛。此外，这些高通滤波器使我们的网络专注于隐写术引入的嵌入伪影，而不是复杂的图像内容。据我们所知，所有经过隐写分析训练的CNN模型都采用了类似的初始化策略[11]、[13]、[22]。

然而，在第一层中使用30个SRM过滤器进行初始化可以作为网络训练的良好起点，但不是最佳终点。根据我们的实验（见表一），在训练期间保持这些过滤器不变通常会导致比更新它们更糟糕的结果。因此，第一卷积层中的所有滤波器都应通过训练与网络中的其他参数一起进行优化。

C. Truncated Linear Unit

动机：激活函数f（·）：R→R将非线性引入到神经网络中，这可以显著地提高特征表示能力。F（·）有多种选择，如传统的Sigmoid和TanH，或最近出现的ReLU（整流线性单位）函数。其中，对于CNN中的卷积层，ReLU是一个值得注意的选择，它可以表示为

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0. \end{cases} \tag{7}$$

ReLU功能已成功应用于CV中的各种任务。对于CV任务，对象分类，目标对象通常可以很容易地从背景中区分出来。换句话说，这些任务中的信号具有高信噪比。在这种情况下，对神经元应用ReLU可以使它们选择性地响应输入中的有用信号，从而产生所谓的稀疏特征。理论和经验论证都表明稀疏表示更可能是线性可分的，并且具有更好的泛化能力[23]。然而，对于我们的隐写分析任务来说，情况完全不同。隐写嵌入过程可以看作是添加低幅度的加性噪声来覆盖图像。与图像内容相比，嵌入信号的幅度要小得多，这意味着信噪比极低。与CV任务或某些高信噪比应用（其中ReLU函数可以很好地适应目标信号的分布）相比，隐写分析中采用的激活函数应考虑嵌入信号的结构，特别是在前几个卷积层中。请注意，在图像隐写术中，嵌入信号通常在[-1,1]的范围内，在我们提出的CNN中引入了一个新的激活函数，称为截断线性单元（TLU），该函数稍微修改了ReLU，定义如下：

$$f(x) = \begin{cases} -T, & x < -T \\ x, & -T \leq x \leq T \\ T, & x > T, \end{cases} \tag{8}$$

where $T > 0$ is the parameter determined by experiments.

在所提出的CNN中，第一卷积层中的加权核被SRM中的基本高通滤波器初始化，这有助于抑制图像内容和提取嵌入信号。在第一次卷积运算的输出中使用TLU有助于：（1）适应嵌入信号的分布（具有低信噪比）；（2）加强CNN，在第一层学习更有效的高通滤波器。根据我们的实验，对于基于CNN的隐写器的其他层，输入信号的分布往往与常规CV任务中的分布更为一致。因此，在这些层中，ReLU函数更可取。

实验验证和分析：我们接着验证TLU在隐写分析中的有效性，并通过实验确定其性能。此外，我们还尝试借助可视化工具解释TLU的功能。

我们基于图2所示的深度卷积模型进行比较。将ReLU作为各层激活函数的CNN即ReLU-CNN训练为基线模型。我们还采用了TLU- CNN，其第一卷积层用TLU代替ReLU。

不同T的ReLU-CNN和TLU-CNN都在0.2 bpp有效载荷下于HILL、S-UNIW ARD和WOW进行训练。实验结果如表2所示。

TABLE II

THE PERFORMANCE OF RELU AND TLU ON RESAMPLED IMAGES IN TERMS OF DETECTION ERROR (P_E) WITH DIFFERENT T SETTINGS. INVOLVED NETWORKS ARE TRAINED ON BOSS+BOWS2+AUG AND TESTED ON BOSS_TEST. THE EMBEDDING PAYLOAD IS 0.2 BPP

Algorithm	ReLU	TLU				
		$T=3$	$T=7$	$T=15$	$T=63$	$T=\infty$
WOW	0.2136	0.1982	**0.1966**	0.2142	0.2139	0.2170
S-UNIWARD	0.2937	**0.2540**	0.2624	0.2653	0.2921	0.2990
HILL	0.2971	**0.2761**	0.2812	0.2894	0.2956	0.2955

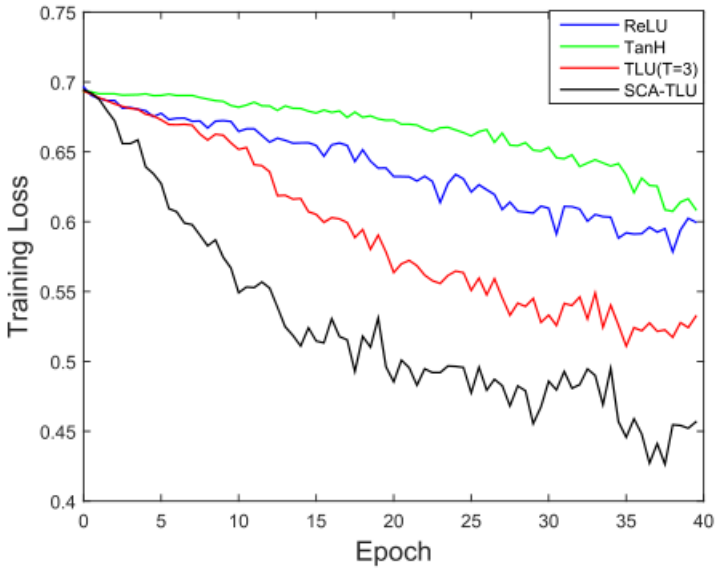
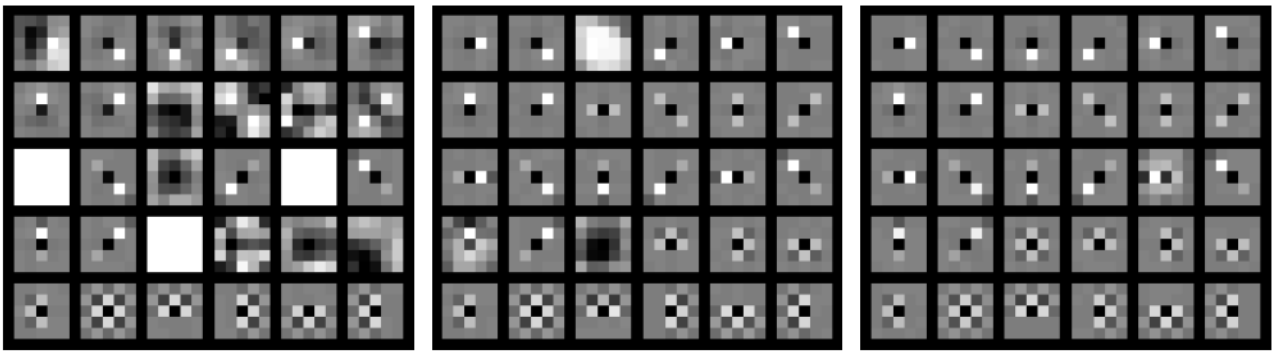


Fig. 3. The convergence performance for training the four involved 10-layer CNN models against S-UNIWARD at 0.2 bpp on resampled BOSS+BOWS2+AUG images. Normalization of the initial high-pass filters is necessary for training with TanH unit. Except for the first layer, all the activation functions are ReLUs.

可以观察到，对于所有三种涉及的隐写方案，TLU-CNN在检测误差方面始终比基线ReLU-CNN在大多数测试参数值方面取得更好的性能，在T=3或T=7时获得最佳性能。随着T值的增加，TLU抑制图像内容的效果逐渐降低，导致性能损失。注意，在极端情况下，当T=∞, TLU成为一个恒等函数（线性激活函数）。有趣的是，即使使用线性激活函数来降低非线性，TLU-CNN仍然具有与ReLU-CNN相当的性能。这可能是由于ReLU的影响，它将所有负输入都设为零，导致嵌入信号中大约50%的信息丢失。

作为一个额外的奖励，它也被观察到TLU-CNN可以训练得比它与ReLU和TanH单位的对手快得多。图3显示了三种网络(t =3的TLU, ReLU和TanH)在训练时的收敛性能。图3显示了S-UNIWARD在0.2 bpp时获得的训练误差随训练图像上epoch数的演化。TLU是根据嵌入信号的分布情况专门设计的。因此，可以更好地利用信息，更有效地训练网络。

为了更好地说明TLU-CNN在隐写分析中的优越性，我们分别在TLU和ReLU训练的第一个卷积层中可视化滤波器。图4（a）和（b）示出了第一卷积层中的滤波器（总共30个滤波器）的可视化，其具有ReLU和TLU（T＝3）。很明显，TLU的采用带来了更多的独特功能和更少的“死”过滤器（与几乎纯白色的功能框相对应的过滤器）。请注意，尽管学习的滤波器与原始SRM滤波器具有相似的形状，但它们实际上具有不同的值。从表I中的结果可以很容易地验证，网络确实可以微调第一层中的SRM滤波器。



(a) (b) (c)

Fig. 4. Visualizations of 1st convolutional layer filters in 3 different models. (a) Filters in ReLU-CNN. (b) Filters in TLU-CNN ($T = 3$). (c) Filters in SCA-TLU-CNN. Using TLU non-linearity and incorporating the knowledge of selection channel can result in more distinctive filters and fewer “dead” filters.

当TLU替换前两层中的RELU时，评估所提出的基于CNN的隐写分析器的性能也是有趣的。如果第一个卷积层中的RELU被TLU替换，而其他层中的RELU保持不变，则该网络称为TLU_n。如表Ⅲ所示，尽管TLU_1和TLU_3的结果相似，但TLU_4及以上的检测精度变得更差。这是因为当网络变深时，TLU输出的分布趋向于与ReLU的分布一致。因此，最好在相对较深的卷积层中使用RELU。考虑到CuDNN库可以加速ReLU的计算，而TLU_3需要更多的训练周期，我们在实现中选择TLU_1作为TLU-CNN模型，以便在训练时间和检测性能之间实现良好的折衷。

D. Incorporating the Knowledge of Selection Channel

E. Curriculum Learning for Low Payload Steganalysis

IV. EXPERIMENTALRESULTS ANDANALYSIS

在这一部分中，我们进行了大量的实验来证明我们提出的CNN模型的可行性和有效性。我们将我们的模型与最先进的手工制作的功能集SRM及其选择通道感知变体maxSRMd2进行比较。为了公平比较，所有涉及的隐写分析方法都在相同的数据集上进行测试。

A. The Steganographic Schemes

在我们的实验中，几种最先进的内容自适应隐写方法在空间域，例如。G使用S-UNIWARD、WOW和HILL来评估相关隐写分析仪的性能。所有嵌入算法都是基于公开的代码在STC模拟器上实现的。注意，在我们的实现中，我们使用Matlab代码中带有随机嵌入键的代码，而不是带有固定嵌入键的C代码版本的模拟器工具（S-UNIWARD，WOW）。这是因为我们在实验中发现，尽管基于CNN的隐写分析器可以实现非凡的检测性能（比如，在0.2 bpp的情况下，WOW的检测误差小于0.1），但如果CNN在模拟器生成的数据集上使用固定的嵌入密钥进行训练，其性能将显著降低（检测误差接近0.5）当使用另一个嵌入键生成测试数据集时。换句话说，训练的CNN过度拟合到训练集中的特定嵌入键，根本没有泛化能力。类似的问题也在[22]中报告，作者使用相同的嵌入密钥创建用于训练的隐藏图像。

B. The Datasets and Data Augmentation

TABLE V

DETECTION ERROR (P_E) OF THREE STEGANALYSIS SCHEMES TRAINED ON DIFFERENT DATASETS AND TESTED ON BOSS_TEST OF RESAMPLED IMAGES, FOR WOW AT 0.2 BPP

Algorithms	BOSS	BOSS+BOWS2	BOSS+BOWS2+AUG
SRM	0.3266	**0.3228**	N/A
maxSRMd2	0.2424	**0.2325**	N/A
TLU-CNN	0.3364	0.2693	**0.1982**

TABLE VI

DETECTION ERROR (P_E) OF THREE STEGANALYSIS SCHEMES TRAINED ON DIFFERENT DATASETS AND TESTED ON BOSS_TEST OF CROPPED IMAGES, FOR WOW AT 0.2 BPP

Algorithms	BOSS	BOSS+BOWS2	BOSS+BOWS2+AUG
SRM	0.3865	**0.3853**	N/A
maxSRMd2	**0.3075**	0.3092	N/A
TLU-CNN	0.4205	0.3512	**0.2808**

TABLE VII

DETECTION ERROR (P_E) OF THREE STEGANALYSIS SCHEMES TRAINED ON DIFFERENT DATASETS AND TESTED ON BOSS_TEST OF SUBSAMPLED IMAGES FOR WOW AT 0.8 BPP

Algorithms	BOSS	BOSS+BOWS2	BOSS+BOWS2+AUG
SRM	**0.2300**	0.2332	N/A
maxSRMd2	0.1813	**0.1807**	N/A
TLU-CNN	0.1991	0.1569	**0.1182**

对于基于CNN的隐写分析，最好采用较大的训练集，以避免过拟合。表V-VII总结了我们基于CNN的隐写分析器和其他两个竞争的隐写分析方案在不同训练集上的性能，并分别对重采样、裁剪和下采样的图像进行BOSS_test测试，对0.2或0.8 bpp的WOW的性能。可以观察到，所提出的TLU-CNN在BOSS上训练时存在大量过拟合。将训练集替换为BOSS+BOWS2后，其性能得到一定程度的提高。使用BOSS+BOWS2+AUG进行网络训练，可以获得最佳的性能。然而，对于所涉及的手工特性集，例如SRM和maxSRMd2，情况就不同了。对于重新采样的图像，最好的选择是BOSS+BOWS2。而对于裁剪和下采样图像，BOSS和BOSS+BOWS2的性能没有明显差异。SRM和maxSRMd2在BOSS+BOWS2+AUG上的实验是没有意义的，因为这些特性已经对称了。旋转或镜像的图像将导致集成分类器的基础FLD学习器中的重复特征和奇异矩阵。因此，对于本文采用的图像数据集，为了公平起见，SRM和maxSRMd2的训练集是BOSS+BOWS2，而我们的CNN模型使用的是BOSS+BOWS2+AUG。并利用上述BOSS_test对所有方案的性能进行评估。为了重复实验，我们创建了三个不同的训练集和测试集。每个CNN模型将使用相同的超参数在三个训练集上独立训练，并在它们相关的测试集上进行测试。然后取测试结果的平均值作为模型的最终性能。

C. Implementation Details

我们使用Caffe[29]对所提出的CNN模型进行了必要的修改。值得注意的是，我们使用AdaDelta[30]来训练我们的网络，而不是SGD，因为我们在早期的实验中发现，使用AdaDelta网络可以学习得更快，取得更好的结果。因此，我们所描述的以下参数都是基于AdaDelta的:mini-batch size为32，包含16对cover和stego图像;动量值为0.95，权值衰减为5×10−4;AdaDelta的delta值为1×10−8。在训练期间进行数据增强，并将相同的旋转或镜像操作应用于小批处理中的一对图像。使用“Xavier”初始化[31]初始化第2 - 9层权重，初始偏差设置为0.2。最后一层全连接层初始化时使用均值为零、标准差为0.01的高斯源获得的随机值，初始偏差设置为零。在上述设置的基础上，对网络进行训练，使交叉熵损失最小化。

在训练过程中，我们使用Caffe的“多步”策略来调整学习率。当训练迭代等于其中一个指定的步长值时，学习率除以5。以重采样图像上0.2 bpp的TLU-CNN for WOW为例，当初始值为0.4时，在50万、60万和65万次迭代时，学习率分别降低到0.08、0.016和0.0032。1注意，不同有效载荷下不同的嵌入方案，我们实际上是针对不同难度的任务训练cnn，这意味着需要使用不同的配置来控制学习率。由于空间限制，

对于我们来说，给出所涉及的每个CNN模型的所有步长值是不切实际的。作为一种替代方法，我们将详细阐述如何确定这些步骤值的规则。解决这一问题的关键在于在训练过程中对验证集的误差和准确性进行监控。当误差不降低，准确率不增加时，应改变学习率。[9]和[32]也采取了类似的政策。注意，对于每个模型，我们使用第IV-B节中描述的方法创建三个不同的训练/验证/测试集，并根据上面描述的规则从第一个训练/验证集中为这个模型选择步骤值。在其他两个训练集上进行训练时使用相同的步长值。

对于负载为0.2 ~ 0.5 bpp的重采样和裁剪图像以及0.8 bpp的下采样图像所对应的TLU-CNN模型，网络参数从头开始训练，并以0.4的初始学习率在100个epoch时停止。应用课程学习策略训练低嵌入率的其他隐写图像对应的模型，即有效载荷为0.05 ~ 0.1 bpp时的重采样和裁剪图像，有效载荷为0.05 ~ 0.5 bpp时的下采样图像。网络与之前训练的网络进行了微调，初始学习率为0.05。所有的微调程序将在35个epoch停止，除了在0.5 bpp的下采样图像的训练，将在70个epoch停止。

D. Comparison With the State-of-the-Art Steganalyzers in Spatial Domain

在本小节中，我们比较了提出的基于cnn的模型与两种最先进的空间域隐藏分析器的性能，即SRM和maxSRMd2，用于各种有效载荷。表VIII-X显示了所有测试方案在重采样、裁剪和下采样图像检测误差(PE)方面的性能比较。在图5到图7中，我们以函数的形式进一步说明了三种最先进的空间域隐写方案，即WOW, S-UNIW ARD和HILL的检测误差的有效载荷(范围从0.05 BPP到0.5 BPP)，所有涉及的隐写分析器在所有3个图像数据集。

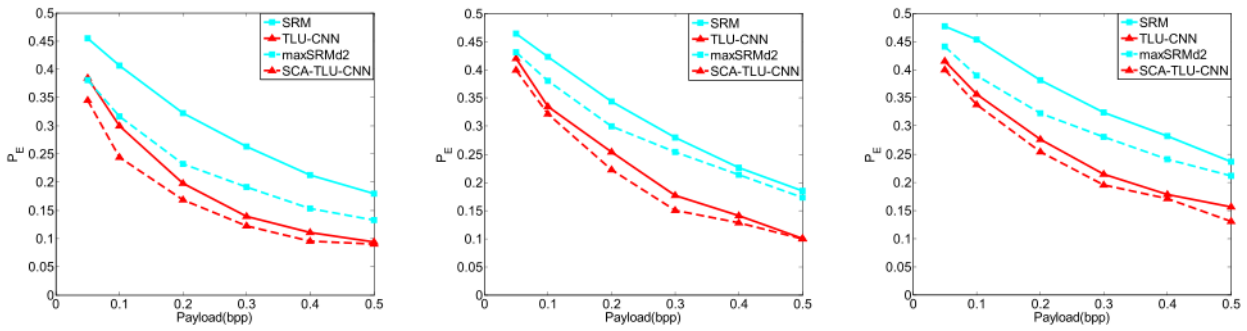
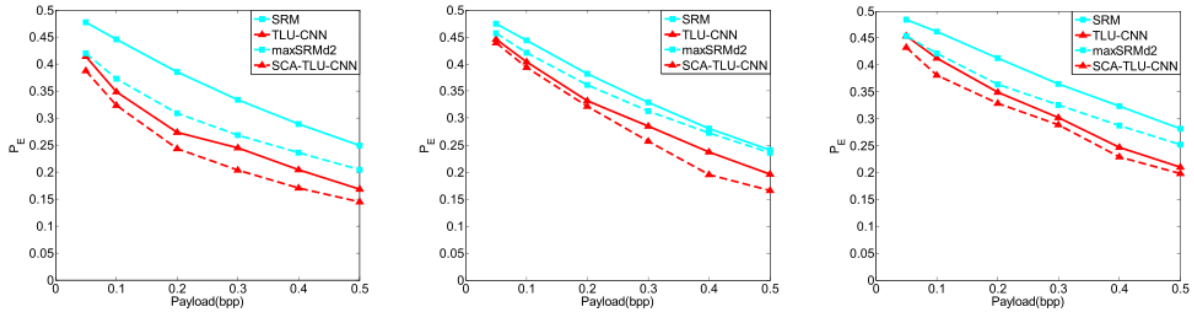
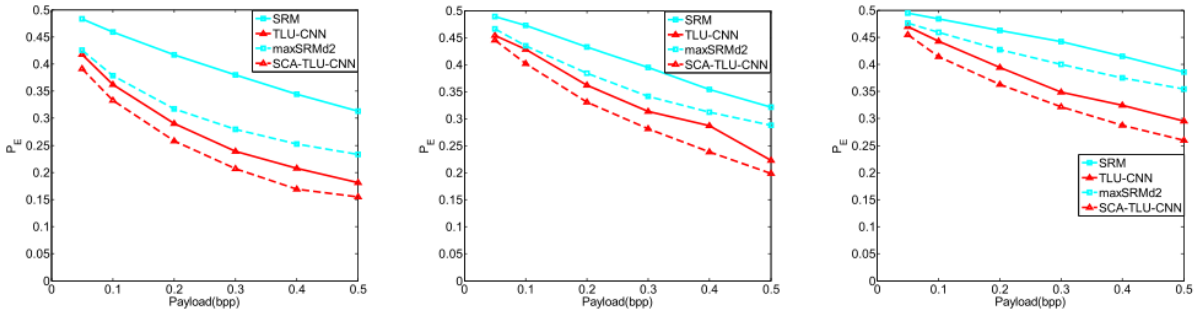


Fig. 5. Detection errors P_E of 3 state-of-the-art steganographic schemes as a function of payload for the involved steganalysis methods.Images are resized to 256 × 256. (a) WOW. (b) S-UNIWARD. (c) HILL.



(a) (b) (c)

Fig. 6. Detection errors P_E of 3 state-of-the-art steganographic schemes as a function of payload for the involved steganalysis methods.Images are cropped into 256 × 256. (a) WOW. (b) S-UNIWARD. (c) HILL.



(a) (b) (c)

Fig. 7. Detection errors P_E of 3 state-of-the-art steganographic schemes as a function of payload for the involved steganalysis methods. Images are subsampled to 256 × 256. (a) WOW. (b) S-UNIWARD. (c) HILL.

TABLE VIII

PERFORMANCE COMPARISON OF THE INVOLVED STEGANALYZERS IN TERMS OF DETECTION ERROR (P_E) FOR 3 STATE-OF-THE-ART STEGANOGRAPHIC SCHEMES AT DIFFERENT PAYLOADS ON RESAMPLED IMAGES

Algorithm	Payload (bpp)	SRM (P_E)	TLU-CNN (P_E)	maxSRMd2 (P_E)	SCA-TLU-CNN (P_E)
WOW	0.05	0.4551	0.3850	0.3810	**0.3450**
	0.1	0.4066	0.3000	0.3163	**0.2442**
	0.2	0.3228	0.1982	0.2325	**0.1691**
	0.3	0.2633	0.1394	0.1918	**0.1229**
	0.4	0.2127	0.1109	0.1536	**0.0959**
	0.5	0.1800	0.0938	0.1331	**0.0906**
S-UNIWARD	0.05	0.4641	0.4200	0.4316	**0.4000**
	0.1	0.4232	0.3350	0.3806	**0.3220**
	0.2	0.3437	0.2540	0.2999	**0.2224**
	0.3	0.2798	0.1772	0.2542	**0.1502**
	0.4	0.2260	0.1410	0.2136	**0.1281**
	0.5	0.1848	0.1003	0.1732	**0.1000**
HILL	0.05	0.4765	0.4150	0.4409	**0.4000**
	0.1	0.453	0.3560	0.3894	**0.3380**
	0.2	0.3811	0.2761	0.3226	**0.2538**
	0.3	0.3236	0.2145	0.2804	**0.1949**
	0.4	0.2818	0.1782	0.2410	**0.1708**
	0.5	0.2363	0.1561	0.2115	**0.1305**

TABLE IX

PERFORMANCE COMPARISON OF THE INVOLVED STEGANALYZERS IN TERMS OF DETECTION ERROR (P_E) FOR 3 STATE-OF-THE-ART STEGANOGRAPHIC SCHEMES AT DIFFERENT PAYLOADS ON CROPPED IMAGES

Algorithm	Payload (bpp)	SRM (P_E)	TLU-CNN (P_E)	maxSRMd2 (P_E)	SCA-TLU-CNN (P_E)

WOW	0.05	0.4772	0.4139	0.4199	**0.3874**
	0.1	0.4460	0.3488	0.3730	**0.3240**
	0.2	0.3853	0.2808	0.3092	**0.2435**
	0.3	0.3337	0.2450	0.2686	**0.2036**
	0.4	0.2887	0.2044	0.2361	**0.1707**
	0.5	0.2496	0.1680	0.2041	**0.1445**
S-UNIWARD	0.05	0.4750	0.4460	0.4571	**0.4390**
	0.1	0.4439	0.4040	0.4206	**0.3938**
	0.2	0.3823	0.3318	0.3614	**0.3218**
	0.3	0.3287	0.2850	0.3132	**0.2571**
	0.4	0.2805	0.2374	0.2721	**0.1955**
	0.5	0.2411	0.1959	0.2355	**0.1660**
HILL	0.05	0.4845	0.4540	0.4536	**0.4325**
	0.1	0.4618	0.4129	0.4211	**0.3806**
	0.2	0.4129	0.3494	0.3638	**0.3288**
	0.3	0.3645	0.3018	0.3253	**0.2885**
	0.4	0.3236	0.2470	0.2874	**0.2291**
	0.5	0.2810	0.2100	0.2520	**0.1977**

TABLE X

PERFORMANCE COMPARISON OF THE INVOLVED STEGANALYZERS IN TERMS OF DETECTION ERROR (P_E) FOR 3 STATE-OF-THE-ART STEGANOGRAPHIC SCHEMES AT DIFFERENT PAYLOADS ON SUBSAMPLED IMAGES

Algorithm	Payload (bpp)	SRM (P_E)	TLU-CNN (P_E)	maxSRMd2 (P_E)	SCA-TLU-CNN (P_E)
WOW	0.05	0.4831	0.4176	0.4254	**0.3916**
	0.1	0.4592	0.3622	0.3788	**0.3333**
	0.2	0.4171	0.2900	0.3176	**0.2585**
	0.3	0.3797	0.2391	0.2796	**0.2070**
	0.4	0.3443	0.2077	0.2523	**0.1691**
	0.5	0.3132	0.1812	0.2335	**0.1547**
S-UNIWARD	0.05	0.4893	0.4541	0.4662	**0.4452**
	0.1	0.4722	0.4283	0.4347	**0.4020**
	0.2	0.4323	0.3618	0.3842	**0.3307**
	0.3	0.3949	0.3137	0.3416	**0.2814**
	0.4	0.3544	0.2872	0.3120	**0.2387**
	0.5	0.3213	0.2226	0.2881	**0.1988**
HILL	0.05	0.4948	0.4697	0.4761	**0.4551**
	0.1	0.4840	0.4430	0.4592	**0.4140**
	0.2	0.4629	0.3940	0.4269	**0.3632**
	0.3	0.4416	0.3490	0.3998	**0.3216**
	0.4	0.4146	0.3245	0.3747	**0.2877**
	0.5	0.3859	0.2950	0.3541	**0.2596**

从图5到图7可以看出，无论采用何种嵌入方法、有效载荷和图像数据集(重采样、裁剪和下采样)，本文提出的TLU-CNN和SCA-TLU-CNN始终以明显的优势优于其他两个手工制作的富模型。一方面，在不考虑选择信道知识的情况下，TLU-CNN模型对于所涉及的嵌入方案、测试有效载荷和图像数据集都比SRM模型获得了显著的性能增益。这对于重新采样和下采样的图像尤其明显。如图5(a)所示，当有效载荷为0.2 bpp时，TLU-CNN相对于SRM，在重采样图像上降低了12.46%的WOW检测误差。从图6还可以看出，对于裁剪后的图像，TLU-CNN的性能增益有所下降。这是因为裁剪后的中心图像通常是原始图像中最复杂的区域，这使得基于CNN和手工制作的隐写分析器更难检测隐写图像。另一方面,对于那些selection-channel-aware计划,我们的SCA-TLU-CNN模型也令人信服地优于maxSRMd2如图5所示,图7的性能差距变得最为明显S-UNIWARD 0.3 bpp图像重新取样,检测误差下降了10.4%。我们认为，使用SRM中的高通滤波器进行正则化初始化、使用TLU非线性以及CNN模型的统一优化框架，大大有助于基于CNN的阶段分析器优于传统启发式特征集的性能。

值得注意的是，尽管TLU-CNN没有明确利用选择信道，但在大多数情况下，它仍然会击败选择信道感知maxSRMd2算法。这一令人惊讶的结果表明，如果TLU-CNN在足够大且多样的训练集上进行训练，它能够隐式地学习特定嵌入方案的选择通道分布。这可能解释了为什么SCA-TLU-CNN的性能没有超过其非选择通道感知版本TLU-CNN。还可以观察到，尽管SCA-TLU-CNN始终比TLU-CNN工作得更好，但在高有效载荷下，性能往往越来越相似，尤其是在重采样图像上的WOW和S-UNIWARD，如图5（a）-（b）所示。这是因为，随着数据有效负载的增加，WOW和S-UNIWARD都变得不那么自适应，因此，SCA-TLU-CNN无法在选择信道的知识方面体现其优势。然而，对于HILL而言，通过一系列过滤操作，它可以在不同的负载下表现出一些“自适应性”，这有助于SCA-TLU-CNN在高数据负载下优于TLU-CNN，尤其是在重采样和次采样图像上。

V. CONCLUSION

现代隐写分析器的范式主要包括残差计算、特征提取和二值分类三个步骤。在本文中，我们提出了一个基于CNN的隐写分析器，它能够很好地模拟和优化这些关键步骤在统一的网络架构。本文提出的CNN与用于CV任务的CNN结构有很大的不同，能够在空间域检测多种不同有效载荷的最先进的隐写方案，并且具有较高的准确性。本文所提出的CNN的第一层权值不是采用随机策略，而是使用SRM中残差映射计算所用的基本高通滤波器初始化，这有助于找到更好的局部最小值作为正则化器。考虑到嵌入信号通常信噪比极低，我们的CNN模型采用了一组混合激活函数，其中除了传统的ReLU函数外，为了更好地适应嵌入信号的分布，在网络的前几层引入了截断线性单元(TLU)函数。最后，通过引入选择信道的知识，进一步提高了所提出的CNN的性能。大量的实验已经进行，表明提出的基于CNN的隐写分析器的性能优于其他最先进的隐写分析方法。