Efficient feature learning and multi-size image steganalysis based on CNN【Zhu-Net基于高效特征学习与多尺度图像隐写分析】

原创

CV误会了我 于 2021-11-01 13:16:43 发布 239 收藏 1

文章标签： cnn 计算机视觉 机器学习

摘要

对于隐写分析，许多研究表明卷积神经网络比传统机器学习方法的两部分结构具有更好的性能。然而，仍然有两个问题需要解决：降低隐写分析特征映射的信噪比和对任意大小的图像进行隐写分析。一些算法需要固定大小的图像作为输入，并且由于未充分利用由各种类型的滤波器获得的噪声残差，因此精度较低。本文针对上述问题，设计了一种基于CNN的改进网络结构。首先，我们使用3×3核代替传统的5×5核，并在预处理层优化卷积核。较小的卷积核用于减少参数数量，并在较小的局部区域对特征进行建模。接下来，我们使用可分离卷积来利用残差的信道相关性，压缩图像内容并增加信噪比（隐藏信号和图像信号之间）。然后，我们使用空间金字塔池（SPP）来聚集局部特征，增强特征的表示能力，并对任意大小的图像进行隐写分析。最后，采用数据扩充技术进一步提高网络性能。实验结果表明，当用于检测WOW和S-UNIWARAD等两种空间算法时，所提出的CNN结构明显优于SRM、Ye-Net、Xu-Net和Yedroudj-Net等其他四种方法。

I. INTRODUCTION

长期以来，隐写术和隐写分析一直是在斗争中发展起来的。隐写术旨在将秘密信息尽可能多地隐藏到特定的cover中，并使cover的变化尽可能少，以便隐写术在视觉质量和统计特征方面接近cover[1,2,3]。同时，隐写分析利用信号处理和机器学习理论，分析了stego和cover的统计差异。它通过增加特征数量和提高分类器性能来提高检测精度[4]。

目前，现有的隐写分析方法包括特定隐写分析算法和通用隐写分析算法。早期的隐写分析方法旨在检测特定的隐写算法[5]，而通用隐写分析算法通常使用统计特征和机器学习[6]。常用的统计特征包括二进制相似性度量特征[7]、DCT特征[8,9]和小波系数特征[10]、共生矩阵特征[11]等。近年来，基于相邻像素间相关性的高阶统计特征已成为隐写分析的主流。这些特征通过捕获与图像隐写术相关的复杂统计特征（如SPAM[12]、丰富模型[13]及其几个变体[14,15]）来提高检测性能。然而，这些先进的方法是基于丰富的模型，包括数以万计的功能。处理这些高维特征将不可避免地导致训练时间增加、过度拟合等问题。此外，基于特征的隐写器能否成功地检测出隐藏的细微变化，很大程度上取决于特征的构造。特征构建需要大量的人工干预和专业知识。

得益于深度学习的发展，卷积神经网络（CNN）在各种隐写分析检测器中表现良好[16,17,18,19]。CNN可以自动从图像中提取复杂的统计相关性，提高检测精度。考虑到GPU内存的限制，现有的隐写术分析器通常针对相对较小的图像（通常256×256）进行训练。但现实世界的图像是任意大小的。这导致了一个问题，即如何使用基于CNN的检测器在固定大小的输入下对任意大小的图像进行隐写。在传统的计算机视觉任务中，输入图像的大小通常直接调整到所需的大小。但是，对于隐写分析来说，这不是一个好的实践，因为像素之间的关系非常弱且相互独立。分类前调整大小会影响探测器的准确性。

在本文中，我们提出了一种新的CNN网络结构“zhu-Net”，以提高空域隐写分析的准确性。所提出的CNN在检测精度和兼容性方面都表现良好，并且与其他CNN相比显示出一些独特的特征，总结如下：

（1） 在预处理层，我们修改了卷积核的大小，并使用SRM[13]的30个基本过滤器在预处理层初始化核，以减少参数数量并优化局部特征。然后，通过训练对卷积核进行优化，以获得更好的精度，加快网络的收敛速度 。
（2）我们使用两个可分离的卷积块来代替传统的卷积层。可分离卷积可用于提取残差的空间相关性和信道相关性，提高信噪比，明显提高精度 。
（3）我们使用空间金字塔池[20]来处理拟议网络中任意大小的图像。空间金字塔池可以将特征映射映射到固定长度，并通过多级池提取特征。

我们设计了实验，将提出的CNN网络与XuNet[17]、YeNet[19]和YedjNet[21]进行比较。提出的CNN显示了卓越的检测精度，甚至超过了最先进的手动功能集，如SRM[13]。

论文的其余部分组织如下。在第二节中，我们简要回顾了空间域中基于卷积神经网络（CNN）的流行图像隐写分析方法的框架。第三节介绍了拟议的CNN，第四节介绍了实验结果和分析。最后，第五节给出了结论。

II. RELATEDWORKS

改进隐写分析CNN结构的常用方法包括：使用截断的线性单元，通过模仿丰富模型提取过程修改拓扑，以及使用更深层次的网络，如ResNet[22]，DenseNet[23]和其他。

Tan等人使用具有四个卷积层的CNN网络进行图像隐写分析[24]。他们的实验表明，具有随机初始化权重的CNN通常无法收敛，使用KV核初始化第一层权重可以提高精度。Qian等人[25]提出了一个隐写分析模型，该模型使用标准CNN结构和高斯激活函数，并进一步证明了转移学习有助于CNN模型检测低有效负载的隐写算法。这些方案的性能与垃圾邮件方案相当或优于SPAM[12]，但仍比SRM方案差[13]。Xu等人[17]提出了一种CNN结构，其中包含一些用于图像分类的技术，如批量归一化（BN）[26]、11个卷积和全局平均池。他们还使用高通滤波器进行预处理，并使用绝对（ABS）激活层。他们的实验显示了更好的性能。通过改进XuCNN，他们实现了更稳定的性能[27]。

在JPEG领域，Xu等人[18]提出了一种基于解压缩图像的网络，与JPEG领域的传统方法相比，该网络具有更好的检测精度。Fridrich等人[28]通过模拟手工特征的传统隐写分析方案，提出了一种具有直方图层的CNN结构，该结构由一组高斯激活函数组成。Ye等人[19]提出了一种CNN结构，其中包含一组高通滤波器用于预处理，并采用了一组混合激活函数来更好地捕获嵌入信号。在选择信道知识和数据扩充的帮助下，他们的模型比经典SRM获得了显著的性能改进。Fridrich[29]提出了一种不同的网络结构，通过手动特征提取来处理任意大小的隐写图像。他们的方案将特征映射的统计元素输入到完全连接的网络分类器。

一般来说，现有网络有两个缺点。
（1） CNN由两部分组成：卷积层和完全连接层（忽略池层等）。卷积层的功能是对输入进行卷积，并输出相应的特征映射。卷积层的输入不需要固定大小的图像，但其输出特征映射可以是任意大小。完全连接的层需要固定大小的输入。因此，完全连接层导致网络的固定大小约束。现有的两种解决方案如下。
1、将输入图像直接调整为所需大小。然而，在隐写分析任务中，图像像素之间的关系是脆弱和独立的。检测隐写术嵌入变化的存在实际上意味着检测添加到cover图像的非常微弱的噪声信号。因此，在将图像输入CNN之前直接调整图像大小将极大地影响网络的检测性能。

2、使用全卷积神经网络（FCN），因为卷积层不需要固定的图像大小。

在本文中，我们提出了第三种解决方案：将特征映射映射到固定大小，然后再将其发送到全连接层，如SPP网络[20]。该网络利用spp模块将特征映射映射到固定长度，实现任意大小图像的隐写分析。

（2） 基于CNN的隐写分析的准确性严重依赖于特征图的信噪比。CNN网络喜欢使用高信噪比来检测stego信号和cover信号之间的微小差异。许多隐写分析方法通常提取图像的残差来提高信噪比。然而，现有的一些方案直接卷积提取的残差，没有考虑残差的跨信道相关性，这并没有很好地利用残差。

在本文中，我们通过以下三种方式提高信噪比。通过减小核大小和提出的“前后梯度下降”方法优化卷积核，利用群卷积分别处理残差的空间相关和信道相关。将这两种方法结合起来，大大提高了隐写分析的准确性。

III. PROPOSED SCHEME

A. Architecture

拟议的基于CNN的隐写分析框架如图1所示。CNN接受大小为256×256的输入图像，并输出两类标签（stego和cover）。提出的CNN由多个层组成，包括一个图像预处理层、两个可分离卷积（sepconv）块、四个用于特征提取的基本块、一个空间金字塔池（SPP）模块和两个完全连接的层以及一个softmax。

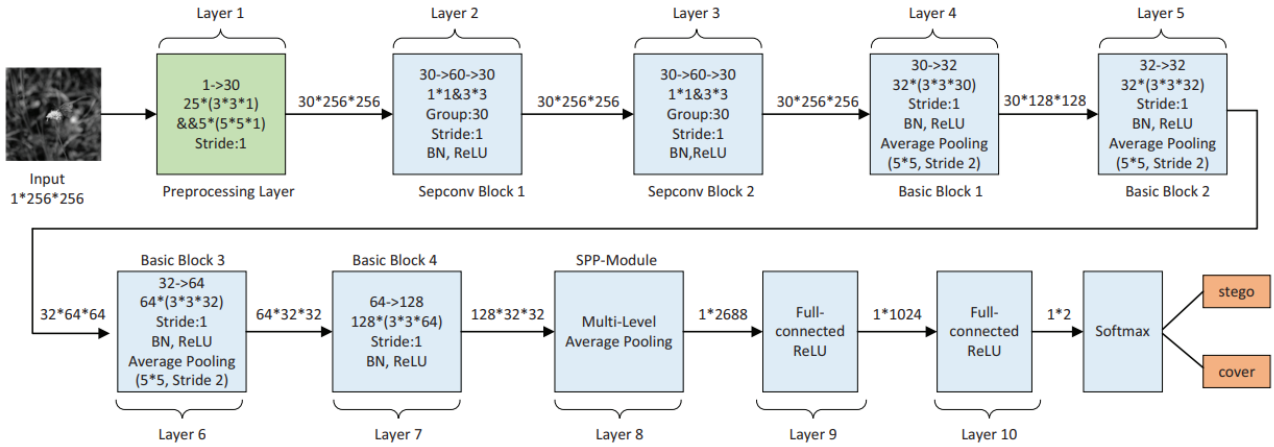


Fig. 1. The architecture of the proposed CNN. For each block, $x_1 \rightarrow x_2; x_2(a*a*x_1)$ denotes the block with the kernel size $a*a$ for x_1 input feature maps and x_2 output feature maps. Batch normalization is abbreviated as BN.

卷积块具有四个标记为“基本块1”到“基本块4”的块，以提取特征图之间的空间相关性，并最终传输到全连接层进行分类。每个基本块由以下步骤组成：

卷积层：与使用大卷积核（例如5×5）的现有网络不同，我们使用小卷积核（例如3×3）来减少参数数量。小卷积核可以增加网络的非线性，从而显著提高特征表示能力。因此，对于基本块1-4，我们将卷积核的大小设置为3×3。对于基本块1到基本块4，有32、32、64、128个通道。每个基本块中的信道数也是基于计算复杂性和网络性能的综合考虑。步幅和填充大小如图1所示。

批标准化（BN）层：批标准化（BN）[26]通常用于在训练期间将每个小批次的分布标准化为零均值和单位方差。使用BN层的优势在于，它有效地防止了深层神经网络中的梯度消失/爆炸和过度拟合[26]，并允许相对较大的学习速率来加速收敛。通过实验，我们发现，没有BN的网络，如YeNet，对参数的初始化非常敏感，并且可能不会在不适当的初始化下收敛。因此，我们在建议的方案中使用BN。

非线性激活函数：对于ZhuNet中的所有块，我们使用经典的整流线性单元（ReLU）作为激活函数，以防止梯度消失/爆炸、生成稀疏特征、加速网络收敛等。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，从而产生更有效的特征。ReLU函数也便于推导，有利于反向传播梯度计算。我们在网络中不使用截断线性单元（TLU），因为我们发现TLU降低了非线性。为了验证这一点，我们将TLU（thresholdT=3）与ReLU进行比较。从表I中，使用ReLU的ZhuNet具有较低的检测各种隐写算法的错误率。如图2所示，ReLU还加快了收敛速度，并显示出比TLU更好的性能。

TABLE I
STEGANALYSIS ERROR RATES COMPARISON OF ZHU-NET WITH TLU AND ZHU-NET WITH RELU AGAINST TWO ALGORITHMS WOW AND S-UNIWARD AT 0.2 BPP AND 0.4 BPP. BOTH NETWORKS ARE TRAINED AND TESTED ON BOSS DATASET.

Algorithms	Zhu-Net with TLU	Zhu-Net with ReLU
WOW(0.2bpp)	0.257	**0.233**
WOW(0.4bpp)	0.138	**0.118**
S-UNIWARD(0.2bpp)	0.316	**0.285**
S-UNIWARD(0.4bpp)	0.188	**0.153**

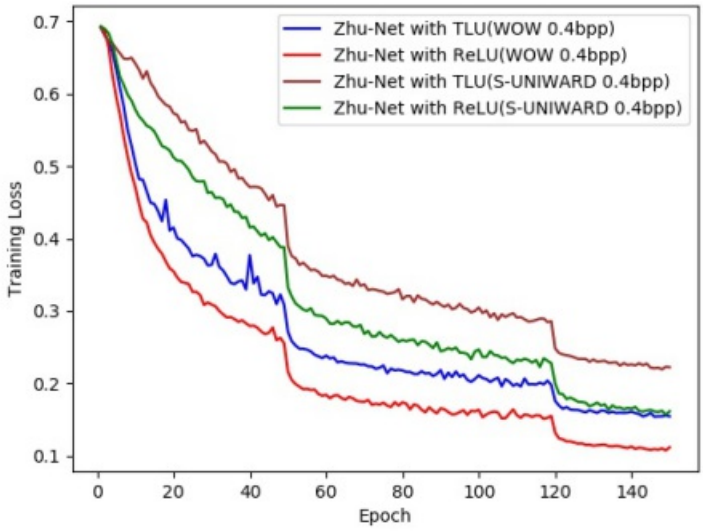


Fig. 2. Comparing convergence performances of training Zhu-Net with TLU and Zhu-Net with ReLU against two algorithms WOW and S-UNIWARD at 0.2 bpp and 0.4 bpp. Both networks are trained and tested on BOSS dataset.

平均池层：在基本块1到基本块3中使用平均池层。它对特征图进行下采样，更好地提取图像特征，缩小特征图的大小，扩大感受野。由于具有不变性，平均池也增强了网络的泛化能力。

注意，我们设计了可分离卷积块来提高信噪比（隐藏信号对图像的信噪比），并有效地从特征中去除图像内容。在最后一块中，我们使用SPP模块来更好地提取特征。SPP模块通过多级池丰富了特征表达式，因此我们的网络可以使用多尺寸图像进行训练和测试。详细内容见第Ⅲ-D节。

B. Improving Kernels

隐写术的嵌入操作可以看作是在覆盖信号中加入一个较小幅度的噪声信号。因此，在网络特征提取之前进行残差计算是一个好主意。在预处理层，我们使用一组高通滤波器（例如，SRM[12]的30个基本高通滤波器，即SPAM滤波器及其旋转对应物，类似于YeNet[19]和Yedrodjnet[21]）从输入图像中提取噪声残差贴图。在[24]中，作者指出，如果没有这样的高通滤波器，CNN的收敛速度应该非常慢。因此，使用多个过滤器可以有效地提高网络性能。我们使用以下策略初始化预处理层的权重。

小尺寸内核：小尺寸卷积内核可以减少参数数量，防止建模较大区域，有效减少计算量。现有的一些方案表明，5×5的核尺寸适合于SRM中的一些滤波器，如“SQUARE5×5”、“EDGE5×5”。但对于剩下的25个滤波器，5×5的协同进化核将模拟一个大的局部区域中的残余元素。因此，我们保留“SQUARE5×5”和“EDGE5×5”，其余25个高通滤波器使用3×3尺寸。我们使用SRM内核初始化卷积内核的中心部分，并将剩余元素填充为零，如图3所示。由这些滤波器计算的残差的两部分被叠加在一起，作为下一个卷积层的输入。

优化内核：建模残差而不是像素值可以提取更健壮的特征。在YEDJNET和Xunet中，预处理层的卷积核在训练过程中是固定的。为了优化利用领域知识设计的SRM手工特征集，我们设计了一种称为“前向-后向梯度下降”的方法，并将其用于预处理层。我们计算残差如下：

For each image $X = X_{ij}$, the residual $R = R_{ij}$ is:

$$R_{ij} = X_{pred}(N_{ij}) - cX_{ij}, \tag{1}$$

where $c \in N$ is the residual order, N_{ij} is the neighboring pixels of X_{ij} and $X_{pred}(.)$ is a predictor of cX_{ij} defined on N_{ij}. In practice, we usually use high-pass filters to achieve $X_{pred}(.)$.

优化内核的完整过程如下所示：

正反向梯度下降法：

正向传播：

输入：图像 X = Xij，高通滤波器k。

输出：噪声残差映射R=Rij。

步骤1：初始化：预处理层的卷积核由SRM中的高通滤波器初始化，卷积核的权重用K表示。

步骤2：计算残差：

$$R = X * K = (\sum_{m,n} X_{i,j}^{m,n} \cdot K^{m,n}), \tag{2}$$

$*$ 表示卷积算子，m表示核K的相应索引。

反向传播：

Back propagation

Input: The gradient of the previous layer δ^{l+1}, the high-pass filter K.

Output: The gradient of the preprocessing layer δ^{l}.

1: Let the backward gradient of the previous layer be δ^{l+1}. Then the gradient of the preprocessing layer is:

$$\delta^{l} = \frac{\partial Loss}{\partial K} = \frac{\partial Loss}{\partial R}\frac{\partial R}{\partial K} = \delta^{l+1} * K, \tag{3}$$

2: Return the gradient of the preprocessing layer δ^{l+1}.

反向传播：

Gradient descent:

Input: The gradient of the preprocessing layer δ^{l}, the high-pass filter K, the learning rate lr.

Output: The optimized kernels $K^{'}$.

1: Optimize the weight of the preprocessing layer by:

$$K^{'} = K - lr * \delta^{l}, \tag{4}$$

2: Return the optimized kernels $K^{'}$.

相应的实验结果如图4和表二所示。我们比较了具有固定核的ZhuNet和具有优化核的网络ZhuNet。从表II可以看出，在检测各种隐写算法时，使用前向-后向优化方法，ZhuNet比使用固定核的网络具有更高的精度。从图4可以看出，我们的网络比固定核的网络收敛更快，训练损失更少。

TABLE II
STEGANALYSIS ERROR RATES COMPARISON BETWEEN ZHU-NET WITH FIXED KERNELS AND ZHU-NET WITH OPTIMIZED KERNELS AGAINST TWO STEGANOGRAPHY ALGORITHMS WOW AND S-UNIWARD AT 0.2 BPP AND 0.4 BPP. BOTH NETWORKS ARE TRAINED AND TESTED ON BOSS DATASET.

Algorithms	Zhu-Net with fixed kernels	Zhu-Net with optimized kernels
WOW(0.2bpp)	0.243	**0.233**
WOW(0.4bpp)	0.130	**0.118**
S-UNIWARD(0.2bpp)	0.324	**0.285**
S-UNIWARD(0.4bpp)	0.169	**0.153**

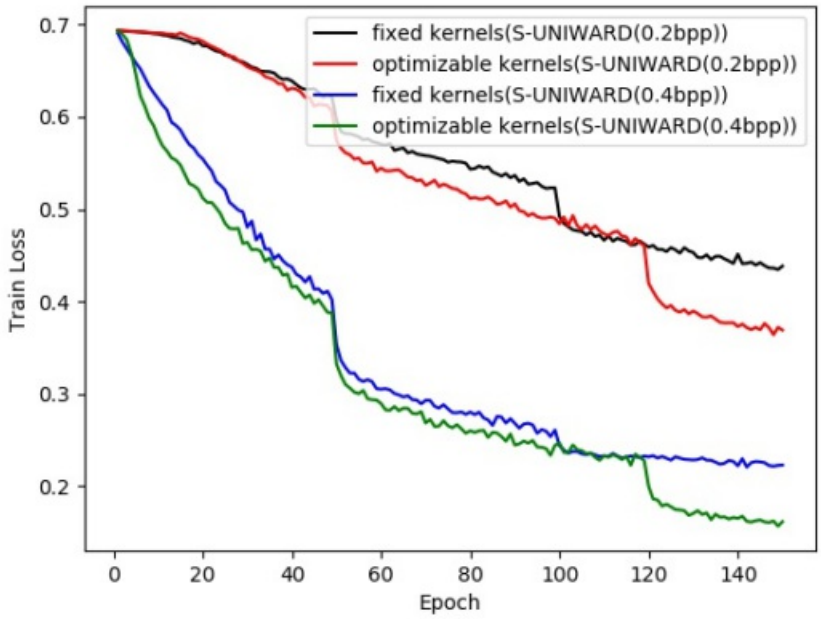


Fig. 4. Comparing convergence performances of training Zhu-Net with fixed kernels and Zhu-Net with optimized kernels against two algorithms WOW and S-UNIWARD at 0.2 bpp and 0.4 bpp. Both networks are trained and tested on BOSS dataset.

C. Separable Convolution

公式问题：现有的隐写分析方案直接在3D空间学习滤波器，而不考虑残差的跨通道相关性，因此残差信息没有得到很好的利用。为了解决这个问题，我们在预处理层后使用了两个可分离的卷积块（即sepconv块），包括1×1卷积和3×3卷积（如图1所示）。

可分离卷积最近在计算机视觉任务中取得了巨大的进展，如初始[30]、例外[31]和其他结构。除此之外，初始模块的变体如图5（a）所示。这个极端版本的inception完全分离了通道之间的相关性，减少了存储空间，增强了模型的表达能力。因此，我们使用异常结构设计相应的sepconv块来实现残差的群卷积。

在我们的方案中，我们假设残差的信道相关性和空间相关性是独立的。sepconv块可以对高通滤波器生成的每个特征映射执行组卷积，这充分利用了剩余信息并从特征中移除图像内容，以提高信噪比。sepconv模块的设计如图5（b）所示。

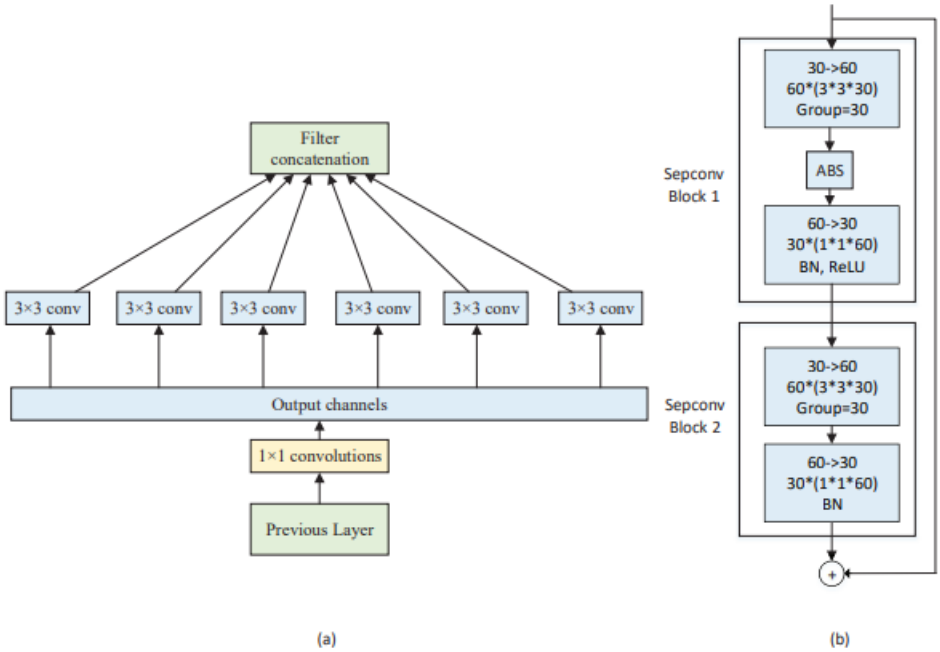


Fig. 5. (a) A variant version of Inception module[31]; (b) structure of sepconv blocks

首先，在sepconv块中执行1×1逐点卷积以提取剩余信道相关性。然后执行3×3深度卷积以提取空间相关性，其中组数为30，sepconv块包括1×1点式卷积和3×3深度卷积。请注意，这两个卷积中没有激活函数。在sepconv块1的11个卷积层之后，考虑到领域知识，我们插入一个绝对激活（ABS）层，使我们的网络学习剩余噪声的符号对称性。我们还在两个sepconv块中使用剩余连接，以加速网络收敛，防止梯度消失/爆炸，并提高分类性能。

实验验证和分析：为了比较ZhuNet和YedroudjNet，我们在第一个卷积层可视化了特征图（当层较深时，CNN的特征图很难解释和可视化）。特征映射很好地描述了特征提取过程。

YedJnet和ZhuNet都使用WOW进行训练，有效载荷为0.2bpp。我们可视化了YedroodJnet的第一个卷积层的特征图和我们的网络的sepconv块2的特征图。stego和cover特征图的比较如图6所示。

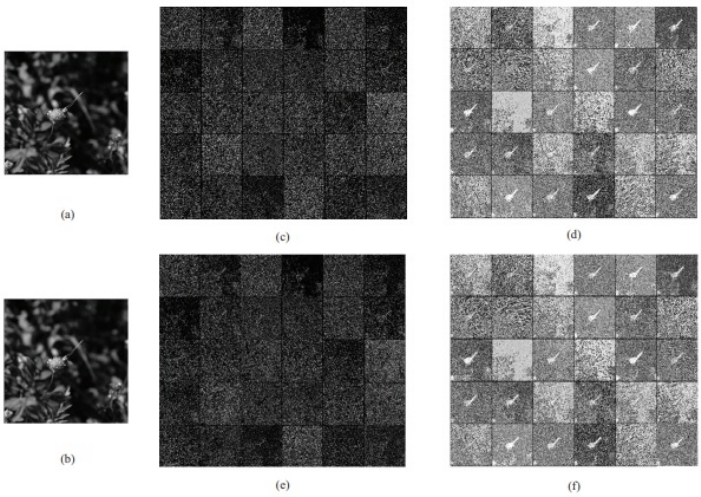


Fig. 6. The comparison of feature maps between Zhu-Net and Yedroudj-Net. (a) Cover image. (b) Stego image. (c) The feature map of cover generated by Zhu-Net. (d) The feature map of cover generated by Yedroudj-Net. (e) The feature map of stego generated by Zhu-Net. (f) The feature map of stego generated by Yedroudj-Net.

实验结果表明，该方法生成的特征映射保留了较少的图像内容信息，并且隐写信号和图像信号之间的信噪比保持在不断提高的水平。无论是cover还是stego，该方法都能提取出具有较强表达能力的特征。同时，每个特征映射之间的相似性相对较低，因此sepconv块便于后续卷积和分类。相比之下，YedroudjNet生成的特征图保留了更多的图像内容，特征图之间的差异不明显。

此外，我们还比较了ZhuNet和YedroudjNet的检测错误率。表三显示了这两个CNN网络相对于SUNIWARD和WOW等两种隐写术方案的性能。实验结果表明，与YENET相比，ZhuNet明显取得了更好的性能，将检测错误率降低2.3%-8. 2%.

TABLE III
STEGANALYSIS ERROR RATES COMPARISON BETWEEN YEDROUDJ-NET AND ZHU-NET AGAINST TWO STEGANOGRAPHY ALGORITHMS WOW AND S-UNIWARD AT 0.2 BPP AND 0.4 BPP. BOTH NETWORKS ARE TRAINED AND TESTED ON BOSS DATASET.

Algorithms	Yedroudj-Net	Zhu-Net
WOW(0.2bpp)	0.278	**0.233**
WOW(0.4bpp)	0.141	**0.118**
S-UNIWARD(0.2bpp)	0.367	**0.285**
S-UNIWARD(0.4bpp)	0.228	**0.153**

TABLE IV
STEGANALYSIS ERROR RATES COMPARISON USING ZHU-NET WITH DIFFERENT NUMBERS OF SEPCONV BLOCKS AGAINST WOW AT 0.2 BPP AND 0.4 BPP. BOTH NETWORKS ARE TRAINED AND TESTED ON BOSS DATASET.

Algorithms	Zhu-Net with full sepconv blocks	Zhu-Net with two sepconv blocks
WOW(0.2bpp)	0.249	**0.233**
WOW(0.4bpp)	0.152	**0.118**

表IV显示了Zhu-Net和Zhu-Net在全sepconv块情况下对算法WOW的检测错误率性能。实验结果表明，用sepconv块替换所有基本块后，检测精度下降。但在低嵌入率（如0.2bpp）下，全sepconv块的Zhu网的精度仍优于Y-edroudj网。如何在CNN中嵌入更多sepcov块需要后续研究。现在，我们在实现中选择了具有两个sepconv块的网络，以实现良好的检测性能。

D. Spatial pyramid pooling module

对于某些隐写分析网络[18,21]，在最后一个卷积层之后添加一个全局平均池（GAP）层进行下采样，这可以大大降低特征维数。对于图像分类，一般采用GAP代替全连通层，以防止过拟合，降低计算复杂度。这种全局平均操作相当于对整个特征图进行建模，从而导致局部特征的信息丢失。然而，对于隐写分析网络，对局部信息进行建模至关重要。

在我们的网络中，我们使用空间金字塔池（SPP）对局部特征地图进行建模，如图7所示。SPP具有以下属性[20]：

（1）SPP为任何大小的输入输出固定长度的特征。

（2） SPP使用多级池来有效地检测对象变形。

（3） 由于输入是任意大小的，SPP可以对任何比例或大小的图像执行特征聚合。

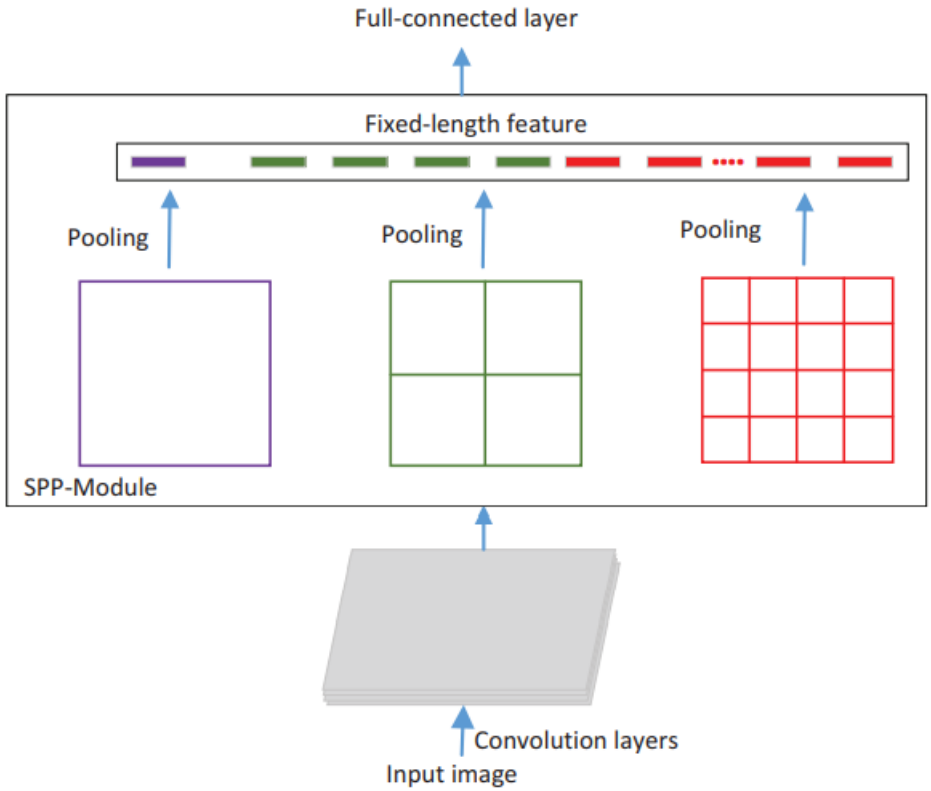


Fig. 7. A network structure with a spatial pyramid pooling layer

与[20]类似，我们将特征映射划分为几个容器。在每个空间单元中，我们汇集每个特征映射的响应（下文使用平均汇集）。空间金字塔池的输出是一个固定的k×M维向量，其中M是单元数，k是最终卷积层中的滤波器数。SPP模块将特征映射到固定长度向量的主要步骤如下所示：

The steps of SPP-module mapping feature maps to fixed length vector

Input: The feature maps after basic block 4 with a size of $a \times a$ and channels of K. an l-level pyramid with $n \times n$ bins in each level.

Output: The fixed length feature with a size of $[1, K \times M]$, where M is the number of bins.

Step 1: For a pyramid level of $n \times n$ bins, implement this pooling level as a sliding window pooling, where the window size win = $\lceil a/n \rceil$, and stride str = $\lfloor a/n \rfloor$ with $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denoting ceiling and floor operations.

Step 2: Implement windows pooling on every feature map, obtain the generated feature with the length of $n \times n$.

Step 3: Repeat step1-step2 for every pyramid level in an l-lever pyramid.

Step 4: Stack all generated feature vectors together (in Pytorch we use torch.cat function). Pre-compute the length of each feature map by $M = \sum_{i=1}^{l} n \times n$, and the total length of feature is $K \times M$.

Step 5: Resize the output feature to a size of $[1, K \times M]$.

我们使用一个三级金字塔池（4×4,2×2,1×1），这意味着箱子的数量是21个（4×42×21×1）。对于给定大小的图像，我们预先计算输出固定长度向量的大小。假设在基本块4之后有一个a×a（例如，32×32）大小的特征映射。当池级别为4×4时，我们将32×32特征图划分为16个小块，即每个小块的大小为8×8。然后在每个8×8块上执行间隙，以获得16维特征向量。在pytorch工具箱中，我们可以使用平均池（strinde:8，kernel:8）来实现这种滑动窗口池操作。2×2和1×1的池化水平相似。最后，我们可以得到一个（4×42×21×1）×k维向量，其中k是最后一个卷积层中的滤波器数。

有趣的是，1×1级池实际上等于许多隐写分析网络中使用的全局平均池层。这表明我们从不同层次的特征图中收集信息，不仅整合了不同尺度的特征，而且更好地模拟了局部特征。

为了验证SPP模块在特征提取中的有效性，我们将ZhuNet（带SPP模块）与YeNet和YedroDjnet（带GAP模块）进行了比较。所有网络均针对WOW和S-UNIWARD进行训练，有效载荷为0。2bpp。实验结果如表五所示。考虑到GPU的计算能力和时间限制，我们构建了一个具有两个预定义大小的训练集：224×224和256×256。我们将所有512×512图像重新采样为256×256图像和224×224图像。为了与现有网络相比，测试图像的大小仍然是256×256。

实验结果表明，ZhuNet比YedroDjnet和YeNet具有更高的精度。也就是说，与单尺寸训练相比，多尺寸训练可以略微提高准确性。我们认为，多尺度训练在一定程度上缓解了过度拟合，多尺度数据集增强了网络的泛化能力。

TABLE V
STEGANALYSIS ERROR PROBABILITY COMPARISON OF ZHU-NET WITH DIFFERENT TRAINING SCHEMES AND YEDROUDJ-NET AGAINST THE TWO ALGORITHMS WOW AND S-UNIWARD AT 0.2BPP. BOTH NETWORKS ARE TRAINED AND TESTED ON BOSS DATASET.

Algorithms	WOW (0.2bpp)	S-UNIWARD (0.2bpp)
Ye-Net	0.331	0.400
Yedroudj-Net	0.278	0.248
Zhu-Net wiht 256-size Tested	**0.234**	**0.281**
Zhu-Net with multi-size Tested	0.241	0.289

此外，我们创建了一个随机大小的测试集，其图像大小范围为[224,256]。ZhuNet对WOW和S-UNIWARD的错误率为0.241和0.289(0.2bpp)

实验结果表明，SPP模型的检测效果优于GAP模型，前者具有更好的特征表达能力。使用SPP模块的另一个优点是，它可以处理任意大小的输入。

IV. EXPERIMENTS

A. The environments

在我们的实验中，我们使用了两种著名的自适应阶段图方法，即。ES-UNIW ARD[3]和WOW[2]，通过随机嵌入密钥的Matlab实现。我们提出的CNN网络与四种流行网络进行了比较：XuNet[17]、YeNet[19]、YEDRODJNET[21]和SRM + EC（代表手工制作的特征集，名为空间丰富模型[13]和集成分类器[32]）。所有五个网络都在相同的数据集上进行测试。所有实验均在Nvidia GTX 1080Ti GPU卡上运行。

B. Datasets

在本文中，我们使用标准数据集来测试所提出的网络的性能。两个标准数据集如下所示：

- the BOSSBase v1.01[33] consisting of 10,000 grey-level images of size 512×512, never compressed, and coming from 7 different cameras.
- the BOWS2[34] consisting of 10,000 grey-level images of size 512×512, never compressed, and whose distribution is close to BOSSBase.

Due to our GPU computing power and time limitation, we do all the experiments on images of 256×256 pixels. The specific training set and test set division will be detailed in Section IV-D.

C. Hyper-parameters

我们采用小批量随机梯度下降（SGD）来训练CNN网络。网络的动量和权重衰减设置为0.9和0.0005。

对于GPU内存限制，训练中的最小批量大小设置为16（8个cover/strgo对）。使用Xavier方法初始化所有层[35]。基于上述设置，训练网络以最小化交叉熵损失。在训练期间，我们调整学习率如下（初始化为0.005）。

当训练迭代等于一个指定的步长值时，学习率将除以5。具体来说，学习率将分别在50、150和250时下降。在以后的训练中，使用较小的学习率可以有效地减少训练损失，提高准确性。CNN的培训长达400个迭代。事实上，我们经常在400个epoch之前停止训练，以防止过度适应。也就是说，当训练集上的交叉熵损失不断减少，但验证集上的检测精度开始下降时，我们停止训练。我们在验证集上选择训练最好的模型。

TABLE VI
STEGANALYSIS ERROR RATES COMPARISON USING YEDROUDJ-NET, XU-NET, YE-NET, AND SRM+EC AGAINST TWO STEGANOGRAPHY ALGORITHMS WOW AND S-UNIWARD AT 0.2 BPP AND 0.4 BPP. ALL NETWORKS ARE TRAINED AND TESTED ON BOSS DATASET.

Algorithms	WOW (0.2bpp)	WOW (0.4bpp)	S-UNIWARD (0.2bpp)	S-UNIWARD (0.4bpp)
SRM+EC	0.365	0.255	0.366	0.247
Xu-Net	0.324	0.207	0.391	0.272
Ye-Net	0.331	0.232	0.400	0.312
Yedroudj-Net	0.278	0.141	0.367	0.228
Zhu-Net	**0.233**	**0.118**	**0.285**	**0.153**

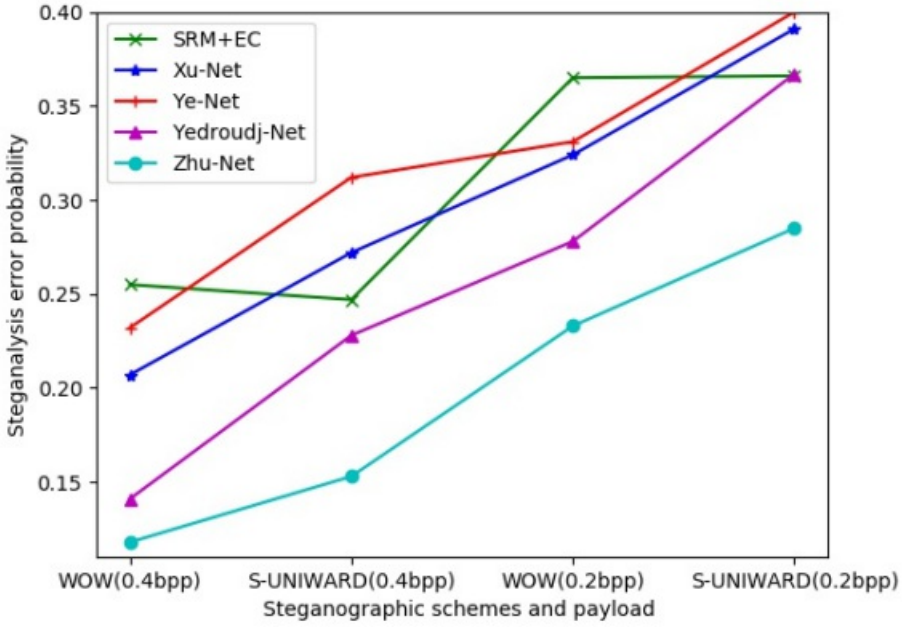


Fig. 8. Steganalysis error rates comparison of the five steganalysis methods against two algorithms WOW and S-UNIWARD at 0.2 bpp and 0.4 bpp. All networks are trained and tested on BOSS dataset.

D. Results

未增加数据的结果：在表六中，我们报告了未增加数据的隐写术分析器的性能比较。BOSSBase图像被随机分成一个包含4000个cover和stego图像对的训练集、一个包含1000个图像对的验证集和一个包含5000个图像对的测试集。为了进行公平比较，我们报告了YEDJNET、YeNet、XuNet和空间丰富模型集成分类器（SRM + EC）在有效载荷为0时相对于嵌入算法WOW和S-UNIWARD的性能。0.2bpp和0.4bpp.

如图8所示，无论采用何种嵌入方法和有效载荷，本文提出的网络的性能都明显优于其他网络。由于CNN的特征提取能力，与传统的SRM+EC网络相比，该网络的错误率降低了8.1% ~ 13.7%。结果还表明，在一个统一的框架内，利用该网络优化特征提取和分类是有效的。

另外，对于不同有效载荷的S-UNIWARD和WOW，所提出的网络比Xu-Net好8.9%～11.9%，比Ye-Net好9.8%～15.9%，比Yedroudj-Net好2.3%～8.2%。结果表明，该网络能有效提取残差的相关性，并具有良好的网络结构，包括spp模块的多级池化，提高了精度。简单地说，实验证明了Zhu-Net在任何有效载荷下对各种隐写方案的性能优于其他网络。注意，上面的实验没有使用迁移学习或数据库虚拟扩展等技巧。

数据增强的结果:通过增加训练数据库的规模，数据增强可以有效地提高网络的性能。使用大型数据库可以提高精度，避免过拟合。但是，传统的数据增强解决方案，如剪切和调整大小是对于隐写分析来说不是一个好的选择，因为这些解决方案会破坏像素的相关性，并大大降低网络的性能。

为了研究增加数据集对性能的影响，我们采用以下数据增强方案。所有的照片被重新采样到256×256像素的大小(使用Matlab中的“imresize()”函数，带有默认设置);

(1)训练集BOSS:将BOSSBase图像随机分为训练集(包含4000张cover和stego图像对)、验证集(包含1000张图像对)和测试集(包含5000张图像对)。
(2)训练集BOSS+BOWS2:在训练集BOSS的基础上，增加10000对cover/stego对(通过对BOWS2Base[36]重采样得到)到训练集。训练数据库现在包含14000对cover/stego图像，验证集包含1000对BOSS图像。
(3)训练集BOSS+BOWS2+DA:通过对BOSS+BOWS2训练集进行保标签翻转和旋转，实际上扩充了数据库BOSS+BOWS2+DA。因此，BOSS+BOWS2训练集的大小增加了8倍，最终得到了由11.2万对遮挡/隐写图像组成的学习数据库。验证集包含来自BOSS的1,000对。
(4)测试集BOSS:包含除训练集BOSS外的其余5000张图像。

TABLE VII
STEGANALYSIS ERROR RATES COMPARISON USING YEDROUDJ-NET, YE-NET AND ZHU-NET ON WOW AT 0.2 BPP WITH A LEARNING BASE AUGMENTED WITH BOWS2, AND DATA AUGMENTATION

Algorithms	BOSS	BOSS+BOWS2	BOSS+BOWS2+DA
Ye-Net	0.331	0.261	0.222
Yedroudj-Net	0.278	0.237	0.208
Zhu-Net	**0.233**	**0.178**	**0.131**

TABLE VIII
STEGANALYSIS ERROR RATES COMPARISON USING YEDROUDJ-NET, YE-NET AND ZHU-NET ON S-UNIWARD AT 0.2 BPP WITH A LEARNING BASE AUGMENTED WITH BOWS2, AND DATA AUGMENTATION

Algorithms	BOSS	BOSS+BOWS2	BOSS+BOWS2+DA
Ye-Net	0.400	-	0.335
Yedroudj-Net	0.366	0.344	0.311
Zhu-Net	**0.285**	**0.243**	**0.171**

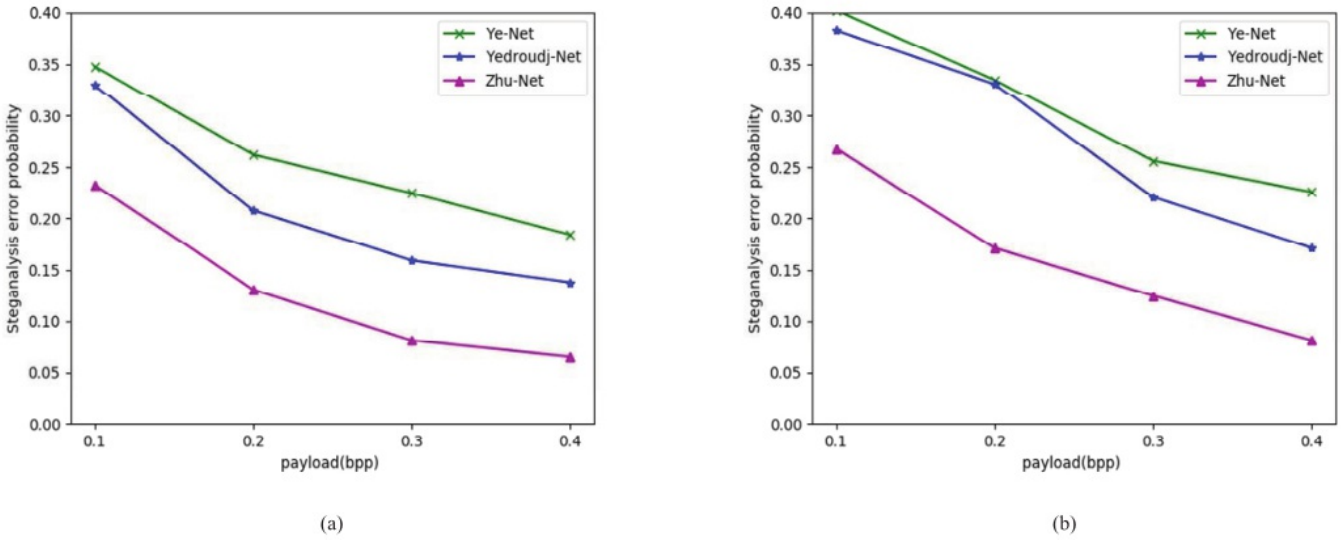
表7和表8为不同训练集上的Yedroudj-Net、Ye-Net和Zhu-Net与有效载荷为0.2 bpp的嵌入算法WOW和SUNIWARD的对比。实验结果表明，随着训练集的增加，与仅使用BOSS训练集相比，所有网络的检测性能都有所提高。对于0.2 bpp的WOW，使用训练集BOSS+BOSW2比仅使用BOSS训练集降低了5.5%的错误率，在所有同类中取得了最好的结果。yedrouj - net和yeenet的错误率分别降低了4.1%和7%。同样，对于S-UNIWARD，在0.2 bpp时，Ye-Net、Yedroudj-Net、Zhu-Net的检测错误率比仅使用BOSS训练数据集分别降低了2.2%和3.6%。Zhu-Net仍然在所有同行中取得了最好的成绩。结果表明，数据增强有效地缓解了过拟合问题。

这促使我们使用更大的数据集进行训练。我们进一步训练了BOSS + BOWS2 + DA三个网络。结果表明，所有基于cnn的方法都提高了性能。与仅使用BOSS的训练相比，ZhuNet的检测误差分别比WOW和SUNIWARD降低了10.2%和11.4% (Ye-Net分别降低了10.9%和6.5%，Yedroudj-Net分别降低了7%和5.5%)。

在表IX和图9中，我们进一步说明了三种基于cnn的隐写分析器在不同载荷下对WOW和S-UNIWARD的检测误差。我们注意到，与其他基于cnn的网络相比，Zhu-Net在不同的数据集和不同的隐写算法上取得了显著的改进和最好的结果。同样，我们将此归功于Zhu-Net良好的网络结构，包括二次conv块和spp模块。

TABLE IX
STEGANALYSIS ERROR RATES COMPARISON USING YEDROUDJ-NET, YE-NET AND ZHU-NET ON WOW AT DIFFERENT PAYLOADS WITH DATA AUGMENTATION

Algorithms	Payload (bpp)	Ye-Net[13]	Yedroudj-Net[31]	Zhu-Net
WOW	0.1	0.348	0.330	**0.233**
	0.2	0.262	0.208	**0.131**
	0.3	0.225	0.189	**0.084**
	0.4	0.184	0.158	**0.065**
S-UNIWARD	0.1	0.400	0.383	**0.268**
	0.2	0.335	0.331	**0.171**
	0.3	0.256	0.221	**0.125**
	0.4	0.226	0.171	**0.081**



(a) (b)

Fig. 9. Steganalysis error rates comparison using YedroudjNet, Ye-Net and Zhu-Net on S-UNIWARD and WOW at different payloads. (a)WOW (b) S-UNIWARD

所有的实验都表明，为了有效地进行特征提取和分类，CNN需要足够的样本进行训练，即使11.2万对图片也可能不够。如何进一步增加数据集以满足隐写分析任务的需要，需要进一步的研究。

V. CONCLUSION

对于隐写分析研究人员来说，使用CNN而不是传统的手工特征——在Rich模型上训练的集成分类器，具有显著的优越性。在本文中，我们设计了一种新的CNN结构用于隐写分析。与现有的基于cnn的网络相比，该网络有了很大的改进。该网络的优点在于:
(1)改进预处理层的卷积核，提取图像残差。更好的卷积核减少了参数和模型局部特征的数量;
(2)利用可分离卷积提取残差的通道相关性和空间相关性，从而去除特征中的图像内容，提高信噪比。利用预处理层残留更有效;
(3)使用spp模块代替全局池层。通过使用不同层次的平均池获取多层次的特征，提高了网络性能。同时，spp模块是处理不同尺寸的灵活解决方案。它可以将特征图映射到固定的维数，可以在不损失任何精度的情况下检测任意大小的图像。
最后，通过使用更大的数据集进一步提高了所提出的CNN的性能。实验结果表明，所提出的CNN网络在检测准确率上明显优于其他网络。