Steganalysis of JPEG Images Using Rich Models

Jan Kodovský and Jessica Fridrich

Department of Electrical and Computer Engineering
Binghamton University, Binghamton, NY 13902-6000, USA

ABSTRACT

In this paper, we propose a rich model of DCT coefficients in a JPEG file for the purpose of detecting steganographic embedding changes. The model is built systematically as a union of smaller submodels formed as joint distributions of DCT coefficients from their frequency and spatial neighborhoods covering a wide range of statistical dependencies. Due to its high dimensionality, we combine the rich model with ensemble classifiers and construct detectors for six modern JPEG domain steganographic schemes: nsF5, model-based steganography, YASS, and schemes that use side information at the embedder in the form of the uncompressed image: MME, BCH, and BCHopt. The resulting performance is contrasted with previously proposed feature sets of both low and high dimensionality. We also investigate the performance of individual submodels when grouped by their type as well as the effect of Cartesian calibration. The proposed rich model delivers superior performance across all tested algorithms and payloads.

在本文中，我们提出了一个丰富的模型的 DCT 系数（依赖关系系数：Discrete Cosine Transform 离散余弦变换是与傅里叶变换相关的一种变换，它与离散傅里叶变换类似，但是只使用实数。）在一个 JPEG 文件，以检测隐写嵌入的变化。该模型是由一组较小的子模型系统地建立起来的，这些子模型是由覆盖了广泛统计相关性的频率和空间邻域的 DCT 系数的联合分布形成的。由于其高维度，我们丰富的模型与系综分类器相结合，构建现代 JPEG 探测器六域隐写方案：nsF5，MBS，YASS，和计划利用边信息嵌入在不压缩的形式：MME，BCH，BCHopt。所得到的性能与之前提出的低维和高维特征集进行了对比。我们还研究了按类型分组时单个子模型的性能以及笛卡尔校准的效果。所提出的富模型在所有经过测试的算法和有效载荷下都具有卓越的性能。

1. MOTIVATION

Modern image-steganography detectors consist of two basic parts: an image model and a machine learning tool that is trained to distinguish between cover and stego images represented in the chosen model. The detection accuracy is primarily determined by the image model, which should be sensitive to steganographic embedding changes and insensitive to the image content. It is also important that it captures as many dependencies among individual image elements (DCT coefficients) as possible to increase the chance that at least some of these dependencies will be disturbed by embedding. By measuring mutual information between coefficient pairs, it has been already pointed out [9] that the strongest dependencies among DCT coefficients are between close spatial-domain (inter-block) and frequency-domain (intra-block) neighbors. This fact was intuitively utilized by numerous researchers in the past, who proposed to represent JPEG images using joint or conditional probability distributions of neighboring coefficient pairs [1,3,14,15,17,22] possibly expanded with their calibrated versions [8, 11]. In [9, 10], the authors pointed out that by merging many such joint distributions (co-occurrence matrices), substantial improvement in detection accuracy can be obtained if combined with machine learning that can handle high model dimensionality and large training sets.

现代图像隐写术检测器由两个基本部分组成：图像模型和经过训练的机器学习工具，用于区分所选模型中表示的覆盖图像和隐写图像（cover、stego）。检测精度主要取决于图像模型，模型对隐写嵌入（0.1、0.2、0.3、0.4、0.5）变化敏感，对图像内容不敏感。同样重要的是，它尽可能多地捕获单个图像元素之间的依赖关系(DCT 系数)，以增加至少一部分依赖关系被嵌入干扰的几率。通过测量系数对之间的互信息，[9]已经指出 DCT 系数之间最强烈的依赖关系是在相邻的**空域(块间)和频域(块内)**之间。过去许多研究人员直观地利用了这一事实，他们提出使用相邻系数对的联合或条件概率分布来表示 JPEG 图像[1,3,14,15,17,22]，这些相邻系数对可能会随着其校准版本的扩展而扩展[8,11]。在[9,10]中，作者指出，通过合并许多这样的联合分布(共现矩阵)，如果结合能够处理高模型维数和大训练集的机器学习，可以显著提高检测精度。

批注 [.皓1]: DCT 系数包含一个 DC 系数和多个 AC 系数。
以 8*8DCT 为例，DCT 系数矩阵第一个数是 DC 系数，其他 63 个数的是 AC 系数。
量化后得到的仍是 64 个系数，量化并没有改变系数的性质。大家知道 DCT 变换是将数据域从时（空）域变换到频域，在频域平面上变换系数是二维频域变量 u 和 v 的函数。对应于 u=0，v=0 的系数，称做直流分量，即 DC 系数，其余 63 个系数称做 AC 系数，即交流分量。
DC 系数：对应于 u=0，v=0 的系数，称做直流分量，即 DC 系数。
AC 系数：其余 63 个系数称做 AC 系数，即交流分量。

In this paper, we propose a complex (rich) model of JPEG images consisting of a large number of individual submodels. The novelty w.r.t. our previous contributions [9,10] is at least three-fold: 1) we view the absolute values of DCT coefficients in a JPEG image as 64 weakly dependent parallel channels and separate the joint statistics by individual DCT modes; 2) to increase the model diversity, we form the same model from differences between absolute values of DCT coefficients; 3) we add integral joint statistics between coefficients from a wider range of values to cover the case when steganographic embedding largely avoids disturbing the first two models. Finally, the joint statistics are symmetrized to compactify the model and to increase its statistical robustness. This philosophy to constructing image models for steganalysis parallels our effort in the spatial domain [4]. We would like to point out that the proposed approach necessitates usage of scalable machine learning, such as the ensemble classifier that was originally described in [9] and then extended to a fully automatized routine in [10].

在本文中，我们提出了一个包含大量独立子模型的复杂(丰富)的 JPEG 图像模型。我们之前的贡献[9,10]至少有三方面：1)我们将 JPEG 图像中的 DCT 系数的绝对值看作 64 个弱相关的并行通道，并通过①单独的 DCT 模式分离联合统计；2)为了增加模型的多样性，我们从 DCT 系数绝对值的差异中②形成相同的模型；3)当隐写嵌入在很大程度上避免了前两个模型的干扰时，我们在较大范围内的系数之间添加积分联合统计量来覆盖这种情况。最后对联合统计量进行了对称化处理，使模型更加紧凑，增强了模型的统计稳健性。这种为隐写分析构建图像模型的理念与我们在空间域[4]中的努力是一致的。我们想指出的是，所提出的方法需要使用可伸缩的机器学习，比如集成分类器，它最初是在[9]中描述的，然后扩展到一个完全自动化的例程中

批注 [.皓2]: 极值点偏移的常规处理方法

The JPEG Rich Model (JRM) is described in detail in Section 2. In Section 3, it is used to steganalyze six modern JPEG-domain steganographic schemes: nsF5 [5]，MBS [19]， YASS [21]， MME [6]，BCH，BCHopt [18]. In combination with an ensemble classifier, the JRM outperforms not only low-dimensional models but also our previously proposed high-dimensional feature sets for JPEG steganalysis – the CC-C300 [9] and CF* [10]. Afterwards, in Section 4, we subject the proposed JRM to analysis and conduct a series of investigative experiments revealing interesting insight and interpretations. The paper is concluded in Section 5.

第 2 节详细描述了 JPEG 富模型(JRM)。

第 3 节中，它用于隐写分析六种现代 JPEG 域隐写方案：nsF5[5]、MBS[19]、YASS[21]、MME[6]、BCH、BCHopt[18]。结合集成分类器，JRM 不仅优于低维模型，而且还优于我们之前提出的 JPEG 隐写分析的高维特征集 CC-C300[9]和 CF[10]。

第 4 节中，我们对提出的 JRM 进行了分析，并进行了一系列的调查实验，揭示了有趣的见解和解释。第五部分是本文的总结。

E-mail: {jan.kodovsky, fridrich}@binghamton.edu; http://dde.binghamton.edu

Media Watermarking, Security, and Forensics 2012, edited by Nasir D. Memon, Adnan M. Alattar,
Edward J. Delp III, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8303,
83030A · © 2012 SPIE-IS&T · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.907495

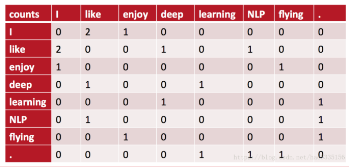
2. RICH MODEL IN JPEG DOMAIN

A JPEG image consists of 64 parallel channels formed by DCT modes which exhibit complex but short-distance dependencies of two types—frequency (intra-block) and spatial (inter-block). The former relates to the relationship among coefficients with similar frequency within the same 8 × 8 block while the latter refers to the relationship across different blocks. Although statistics of neighboring DCT coefficients were used as models in the past, the need to keep the model dimensionality low for the subsequent classifier training usually limited the model scope to co-occurrence matrices constructed from all coefficients in the DCT plane. Thus, despite their very different statistical nature, all DCT modes were treated equally.

一幅 JPEG 图像由 64 个由 DCT 模式组成的并行信道组成（63 个 AC 交流，1 个 DC 直流），DCT 模式表现出复杂但短距离的两种类型的相关性——频率(块内)和空间(块间)。前者是指同一 8×8 块内频率相近的系数之间的关系，后者是指不同块之间的关系。虽然过去使用相邻 DCT 系数①的统计量作为模型，但后续训练分类器时需要保持模型维数较低，因此通常将模型范围限制在由 DCT 平面上所有系数构造的共现矩阵②上。因此，尽管它们（两种类型的相关性）的统计性质非常不同，所有的 DCT 模式都被平等对待。

批注 [.皓3]: 解决词向量相近关系的表示
例如：语料库如下：

- I like deep learning.
- I like NLP.
- I enjoy flying.



counts	I	like	enjoy	deep	learning	NLP	flying	.
I	0	2	1	0	0	0	0	0
like	2	0	0	1	0	1	0	0
enjoy	1	0	0	0	0	0	1	0
deep	0	1	0	0	1	0	0	0
learning	0	0	0	1	0	0	0	1
NLP	0	1	0	0	0	0	0	1
flying	0	0	1	0	0	0	0	1
.	0	0	0	0	1	1	1	0

缺点：面临稀疏性问题、向量维数随着词典大小线性增长。解决：SVD、PCA 降维，但是计算量大

Our proposed rich model consists of several qualitatively different parts. First, in the lines of our previously proposed CF ∗ features, we model individual DCT modes separately, collect many of these submodels a put them together. They will be naturally diverse since they capture dependencies among different DCT coefficients. The second part of the proposed JRM is formed as integral statistics from the whole DCT plane. The increased statistical power enables us to extend the range of co-occurrence features and therefore cover a different spectrum of dependencies than the mode-specific features from the first part. The features of both parts are further diversified by modeling not only DCT coefficients themselves, but also their differences calculated in different directions.

我们提出的富模型由几个性质不同的部分组成。首先，在我们之前建议的 CF 的行中，我们分别对各个 DCT 模式建模，①收集许多子模型并将它们放在一起。由于它们捕获了不同 DCT 系数之间的依赖关系，因此它们自然是不同的。提出的 JRM（JPEG 富模型）的第二部分是由②整个 DCT 平面的积分统计形成的。增加的统计能力使我们能够扩展共现特性的范围，从而覆盖与第一部分中特定于模式的特性不同的依赖范围。通过对 DCT 系数本身的建模，以及对其在不同方向上的差异进行计算，进一步丰富了两部分的特征。

2.1 Notation and definitions（符号和定义）

量化的 DCT 系数的 JPEG 图像的尺寸是 M×N 维会被矩阵 D∈Z$^{M\times N}$表示。$D_{xy}^{(i,j)}$ 表示在 8×8 块中第(i, j)的第(x, y)个 DCT 系数，$(x,y)\in\{0,\ldots,7\}^2, i=1,\ldots,\lceil M/8\rceil$ $j=1,\ldots,\lceil N/8\rceil$. 。或者，我们可以访问单个元素为 D_{ij}， $i=1,\ldots,M,\ j=1,\ldots,N.$ 我们定义如下矩阵:

$$
\begin{aligned}
\mathbf{A}_{i,j}^{\times} &= |\mathbf{D}_{ij}|,\ i=1,\ldots,M,\ j=1,\ldots,N, && (1)\\
\mathbf{A}_{i,j}^{\rightarrow} &= |\mathbf{D}_{ij}|-|\mathbf{D}_{i,j+1}|,\ i=1,\ldots,M,\ j=1,\ldots,N-1, && (2)\\
\mathbf{A}_{i,j}^{\downarrow} &= |\mathbf{D}_{ij}|-|\mathbf{D}_{i+1,j}|,\ i=1,\ldots,M-1,\ j=1,\ldots,N, && (3)\\
\mathbf{A}_{i,j}^{\searrow} &= |\mathbf{D}_{ij}|-|\mathbf{D}_{i+1,j+1}|,\ i=1,\ldots,M-1,\ j=1,\ldots,N-1, && (4)\\
\mathbf{A}_{i,j}^{\rightrightarrows} &= |\mathbf{D}_{ij}|-|\mathbf{D}_{i,j+8}|,\ i=1,\ldots,M,\ j=1,\ldots,N-8, && (5)\\
\mathbf{A}_{i,j}^{\downdownarrows} &= |\mathbf{D}_{ij}|-|\mathbf{D}_{i+8,j}|,\ i=1,\ldots,M-8,\ j=1,\ldots,N. && (6)
\end{aligned}
$$

矩阵 A$^{\times}$由 DCT 系数的绝对值组成，矩阵 A$^{\rightarrow}$,A$^{\downarrow}$,A$^{\searrow}$为块内差异，A$^{\rightrightarrows}$,A$^{\downarrow\downarrow}$是块间差异。所提出的 JRM 的各个子模型将由矩阵系数计算得到的二维共现矩阵构成 A*，⁎∈{×，→，↓，↘，⇉，↓↓}，A*是一个(2T+1)^{2}维的矩阵，其中元素

$$
c_{kl}^{\star}(x,y,\Delta x,\Delta y)=\frac{1}{Z}\sum_{i,j}\left|\left\{\mathbf{T}_{xy}^{(i,j)}\middle|\mathbf{T}=\mathrm{trunc}_T(\mathbf{A}^{\star});\ \mathbf{T}_{xy}^{(i,j)}=k;\ \mathbf{T}_{x+\Delta x,y+\Delta y}^{(i,j)}=l\right\}\right|, \tag{7}
$$

在归一化常数 Z 确保 $\sum_{k,l}c_{kl}=1$，以智能元素截断算子定义为是 trunc$_T$(·)

$$
\mathrm{trunc}_T(x)=\begin{cases}T\cdot\mathrm{sign}(x) & \text{if } |x|>T,\\ x & \text{otherwise.}\end{cases} \tag{8}
$$

在定义(7),我们不限制ΔxΔy 和允许(x+Δx,y +Δy) 超出{0, ···, 7}^{2} 来更容易地描述块间系数对的共现，例如 $T_{x+8,y}^{(i,j)}=T_{xy}^{(i+1,j)}$。

假设主对角线镜像后的自然图像统计量不变，DCT 基函数的对称性关于 8×8 块对角可以替换矩阵 C_T^* 更健壮。

$$\bar{\mathbf{C}}_T^{\times}(x,y,\Delta x,\Delta y) \triangleq \frac{1}{2}\left(\mathbf{C}_T^{\times}(x,y,\Delta x,\Delta y)+\mathbf{C}_T^{\times}(y,x,\Delta y,\Delta x)\right), \quad (9)$$

$$\bar{\mathbf{C}}_T^{\rightarrow}(x,y,\Delta x,\Delta y) \triangleq \frac{1}{2}\left(\mathbf{C}_T^{\rightarrow}(x,y,\Delta x,\Delta y)+\mathbf{C}_T^{\downarrow}(y,x,\Delta y,\Delta x)\right), \quad (10)$$

$$\bar{\mathbf{C}}_T^{\rightrightarrows}(x,y,\Delta x,\Delta y) \triangleq \frac{1}{2}\left(\mathbf{C}_T^{\rightrightarrows}(x,y,\Delta x,\Delta y)+\mathbf{C}_T^{\downdownarrows}(y,x,\Delta y,\Delta x)\right), \quad (11)$$

$$\bar{\mathbf{C}}_T^{\searrow}(x,y,\Delta x,\Delta y) \triangleq \frac{1}{2}\left(\mathbf{C}_T^{\searrow}(x,y,\Delta x,\Delta y)+\mathbf{C}_T^{\searrow}(y,x,\Delta y,\Delta x)\right). \quad (12)$$

因为 $A_{i,j}^{\times}$ 系数是非负的，大部分$\bar{\mathbf{C}}_T^{\times}$是 0，并且它的真实维度只有$(T+1)^2$。焦点是共生的$\bar{\mathbf{C}}_T^{\times}$，$\star\in\{\rightarrow,\searrow,\rightrightarrows\}$，通常非零，但是我们还可以利用它们的符号对称性$(c_{kl}^{\star}\approx c_{-k,-l}^{\star})$，并且定义$\hat{\mathbf{C}}_T^{\star}$

$$\hat{c}_{kl}^{\star}=\frac{1}{2}\left(\bar{c}_{kl}^{\star}+\bar{c}_{-k,-l}^{\star}\right).$$

A 的冗余部分可以去掉，从而得到基于不同的共生的维度的最终形式$\frac{1}{2}(2T+1)^2+\frac{1}{2}$，我们再次定义

$\hat{\mathbf{C}}_T^{\star}(x,y,\Delta x,\Delta y),\ \star\in\{\rightarrow,\searrow,\rightrightarrows\}.$ 这个富模型将仅使用最紧凑的形式构造：$\bar{\mathbf{C}}_T^{\times}$, $\hat{\mathbf{C}}_T^{\rightarrow}$, $\hat{\mathbf{C}}_T^{\searrow}$, and $\hat{\mathbf{C}}_T^{\rightrightarrows}$. 我们注意到，的共现度是由[10]中提出的 F*集演化而来的。不同之处在于，F*在形成协变量之前不接受绝对值。取绝对值降低了维数，使特征更加稳健；它可以看作是另一种对称。在第 3 节中，我们将所提出的 rich 模型的性能与经过笛卡尔校准的 CF*set[10]相比较。

2.2 DCT-mode specific components of JRM（JRM 的特定于 DCT 模式的组件）

根据 DCT 模式的相互位置(x, y)和(x+Δx y+Δy),提取的共线矩阵$\mathbf{C}\in\{\bar{\mathbf{C}}_T^{\times},\hat{\mathbf{C}}_T^{\rightarrow},\hat{\mathbf{C}}_T^{\searrow},\hat{\mathbf{C}}_T^{\rightrightarrows}\}$将被分成 10 个不同性质的子模型 **C(x,y,0,1)意思是 x 是行，y 是列，第三个位置是 0：行不变，第四个位置是 1：列加 1**

1. $\mathcal{G}_{\mathrm{h}}(\mathbf{C}) = \{\mathbf{C}(x,y,0,1)|0\le x;\ 0\le y;\ x+y\le 5\},$
2. $\mathcal{G}_{\mathrm{d}}(\mathbf{C}) = \{\mathbf{C}(x,y,1,1)|0\le x\le y;\ x+y\le 5\}\cup\{\mathbf{C}(x,y,1,-1)|0\le x<y;\ x+y\le 5\},$
3. $\mathcal{G}_{\mathrm{oh}}(\mathbf{C}) = \{\mathbf{C}(x,y,0,2)|0\le x;\ 0\le y;\ x+y\le 4\},$
4. $\mathcal{G}_{\mathrm{x}}(\mathbf{C}) = \{\mathbf{C}(x,y,y-x,x-y)|0\le x<y;\ x+y\le 5\},$
5. $\mathcal{G}_{\mathrm{od}}(\mathbf{C}) = \{\mathbf{C}(x,y,2,2)|0\le x\le y;\ x+y\le 4\}\cup\{\mathbf{C}(x,y,2,-2)|0\le x<y;\ x+y\le 5\},$
6. $\mathcal{G}_{\mathrm{km}}(\mathbf{C}) = \{\mathbf{C}(x,y,-1,2)|1\le x;\ 0\le y;\ x+y\le 5\},$
7. $\mathcal{G}_{\mathrm{ih}}(\mathbf{C}) = \{\mathbf{C}(x,y,0,8)|0\le x;\ 0\le y;\ x+y\le 5\},$
8. $\mathcal{G}_{\mathrm{id}}(\mathbf{C}) = \{\mathbf{C}(x,y,8,8)|0\le x\le y;\ x+y\le 5\},$
9. $\mathcal{G}_{\mathrm{im}}(\mathbf{C}) = \{\mathbf{C}(x,y,-8,8)|0\le x\le y;\ x+y\le 5\},$
10. $\mathcal{G}_{\mathrm{ix}}(\mathbf{C}) = \{\mathbf{C}(x,y,y-x,x-y+8)|0\le x;\ 0\le y;\ x+y\le 5\}.$

前六个子模型捕获块内关系：

1. 水平的（垂直，后对称）邻近对；2. 对角和小对角相邻对；

3. “跳过一个”水平相邻的对；　　　4. 关于 8×8 数据块对对称的位置；

5. “跳过一个”对角和小对角对；　　6. 骑士步位置对

最后四个子模型捕获相邻块的系数之间的块间关系：

7. DCT 模式相同的水平邻居；8. 相同模式下的对角邻居；

9. 同一模式下的小对角邻居；10. 关于 8×8 块的模式对称定位的水平邻居

构成 Gd 和 God 的两个部分被组合在一起，给出了所有子模型大致相同的维度。

自从子模型十组被构造成$\mathbf{C} \in \{\bar{\mathbf{C}}_3^{\times}, \hat{\mathbf{C}}_2^{\rightarrow}, \hat{\mathbf{C}}_2^{\searrow}, \hat{\mathbf{C}}_2^{\rightrightarrows}\}$,共获得 40 个 DCT 模式具体子模型。对于同时出现的 DCT 系数绝对值，我们修正 T=3. 单个矩阵$\bar{\mathbf{C}}_3^{\times}$的维数等于 16。对于基于差异的共现，我们将 T=2 固定为类似的维度 13。较大的 T 值会导致许多未填充的箱子，特别是对于较小的图像。所有引入的子模型的列表，包括它们的总维数，如图 1 所示。

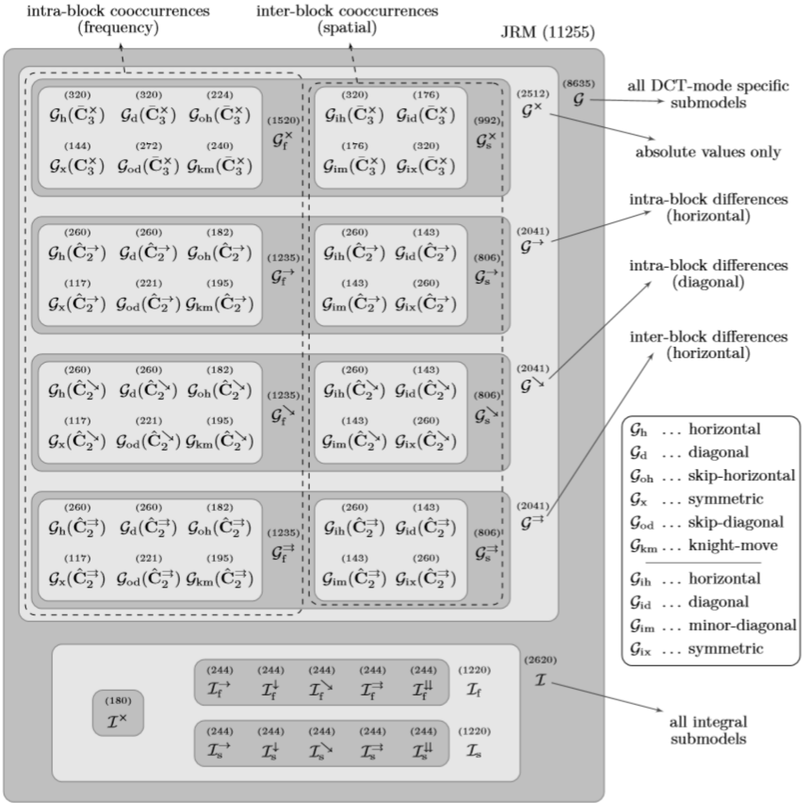

intra-block cooccurrences (frequency)
inter-block cooccurrences (spatial)
JRM (11255)
(8635) $\mathcal{G}$
all DCT-mode specific submodels
(2512) $\mathcal{G}^{\times}$
absolute values only
(1520) $\mathcal{G}_f^{\times}$
(992) $\mathcal{G}_s^{\times}$
(2041) $\mathcal{G}^{\rightarrow}$
intra-block differences (horizontal)
(2041) $\mathcal{G}^{\searrow}$
intra-block differences (diagonal)
(2041) $\mathcal{G}^{\rightrightarrows}$
inter-block differences (horizontal)
(1235) $\mathcal{G}_f^{\rightarrow}$ (806) $\mathcal{G}_s^{\rightarrow}$
(1235) $\mathcal{G}_f^{\searrow}$ (806) $\mathcal{G}_s^{\searrow}$
(1235) $\mathcal{G}_f^{\rightrightarrows}$ (806) $\mathcal{G}_s^{\rightrightarrows}$
(320) $\mathcal{G}_h(\bar{\mathbf{C}}_3^{\times})$ (320) $\mathcal{G}_d(\bar{\mathbf{C}}_3^{\times})$ (224) $\mathcal{G}_{oh}(\bar{\mathbf{C}}_3^{\times})$
(144) $\mathcal{G}_x(\bar{\mathbf{C}}_3^{\times})$ (272) $\mathcal{G}_{od}(\bar{\mathbf{C}}_3^{\times})$ (240) $\mathcal{G}_{km}(\bar{\mathbf{C}}_3^{\times})$
(320) $\mathcal{G}_{ih}(\bar{\mathbf{C}}_3^{\times})$ (176) $\mathcal{G}_{id}(\bar{\mathbf{C}}_3^{\times})$
(176) $\mathcal{G}_{im}(\bar{\mathbf{C}}_3^{\times})$ (320) $\mathcal{G}_{ix}(\bar{\mathbf{C}}_3^{\times})$
(260) $\mathcal{G}_h(\hat{\mathbf{C}}_2^{\rightarrow})$ (260) $\mathcal{G}_d(\hat{\mathbf{C}}_2^{\rightarrow})$ (182) $\mathcal{G}_{oh}(\hat{\mathbf{C}}_2^{\rightarrow})$
(117) $\mathcal{G}_x(\hat{\mathbf{C}}_2^{\rightarrow})$ (221) $\mathcal{G}_{od}(\hat{\mathbf{C}}_2^{\rightarrow})$ (195) $\mathcal{G}_{km}(\hat{\mathbf{C}}_2^{\rightarrow})$
(260) $\mathcal{G}_{ih}(\hat{\mathbf{C}}_2^{\rightarrow})$ (143) $\mathcal{G}_{id}(\hat{\mathbf{C}}_2^{\rightarrow})$
(143) $\mathcal{G}_{im}(\hat{\mathbf{C}}_2^{\rightarrow})$ (260) $\mathcal{G}_{ix}(\hat{\mathbf{C}}_2^{\rightarrow})$
(260) $\mathcal{G}_h(\hat{\mathbf{C}}_2^{\searrow})$ (260) $\mathcal{G}_d(\hat{\mathbf{C}}_2^{\searrow})$ (182) $\mathcal{G}_{oh}(\hat{\mathbf{C}}_2^{\searrow})$
(117) $\mathcal{G}_x(\hat{\mathbf{C}}_2^{\searrow})$ (221) $\mathcal{G}_{od}(\hat{\mathbf{C}}_2^{\searrow})$ (195) $\mathcal{G}_{km}(\hat{\mathbf{C}}_2^{\searrow})$
(260) $\mathcal{G}_{ih}(\hat{\mathbf{C}}_2^{\searrow})$ (143) $\mathcal{G}_{id}(\hat{\mathbf{C}}_2^{\searrow})$
(143) $\mathcal{G}_{im}(\hat{\mathbf{C}}_2^{\searrow})$ (260) $\mathcal{G}_{ix}(\hat{\mathbf{C}}_2^{\searrow})$
(260) $\mathcal{G}_h(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (260) $\mathcal{G}_d(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (182) $\mathcal{G}_{oh}(\hat{\mathbf{C}}_2^{\rightrightarrows})$
(117) $\mathcal{G}_x(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (221) $\mathcal{G}_{od}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (195) $\mathcal{G}_{km}(\hat{\mathbf{C}}_2^{\rightrightarrows})$
(260) $\mathcal{G}_{ih}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (143) $\mathcal{G}_{id}(\hat{\mathbf{C}}_2^{\rightrightarrows})$
(143) $\mathcal{G}_{im}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (260) $\mathcal{G}_{ix}(\hat{\mathbf{C}}_2^{\rightrightarrows})$
$\mathcal{G}_h$ ... horizontal
$\mathcal{G}_d$ ... diagonal
$\mathcal{G}_{oh}$ ... skip-horizontal
$\mathcal{G}_x$ ... symmetric
$\mathcal{G}_{od}$ ... skip-diagonal
$\mathcal{G}_{km}$ ... knight-move
$\mathcal{G}_{ih}$ ... horizontal
$\mathcal{G}_{id}$ ... diagonal
$\mathcal{G}_{im}$ ... minor-diagonal
$\mathcal{G}_{ix}$ ... symmetric
(2620) $\mathcal{I}$
all integral submodels
(180) $\mathcal{I}^{\times}$
(244) $\mathcal{I}_f^{\rightarrow}$ (244) $\mathcal{I}_f^{\downarrow}$ (244) $\mathcal{I}_f^{\searrow}$ (244) $\mathcal{I}_f^{\rightrightarrows}$ (244) $\mathcal{I}_f^{\downdownarrows}$ (1220) $\mathcal{I}_f$
(244) $\mathcal{I}_s^{\rightarrow}$ (244) $\mathcal{I}_s^{\downarrow}$ (244) $\mathcal{I}_s^{\searrow}$ (244) $\mathcal{I}_s^{\rightrightarrows}$ (244) $\mathcal{I}_s^{\downdownarrows}$ (1220) $\mathcal{I}_s$

图 1 所示：提出的 JPEG 富模型及其分解为单独的子组和子模型。数字表示相应集合的维数。笛卡尔校准使所有显示值加倍。

2.3 Integral components of JRM（JRM 的组成部分）

第 2.2 节中介绍的特定于模式（DCT 模式）的子模型为丰富的模型提供了更细的“粒度”，但代价是一次只能利用 DCT 平面的一小部分。为了不丢失整个 DCT 平面的积分统计能力，并覆盖更大范围的 DCT 系数，我们现在通过补充在所有 DCT 模式上积分的额外共现矩阵来确定丰富的模型。我们对 DCT 系数绝对值的共现都这样做$\bar{\mathbf{C}}_T^{\times}$，$\hat{\mathbf{C}}_T^{\star}, \star \in \{\rightarrow, \downarrow, \searrow, \rightrightarrows, \downdownarrows\}$.由于积分箱比 DCT 模式的特定箱更容易填充，所以我们将 T 增加到 5。积分子模型定义如下:

1. $\mathcal{I}^{\times}=\left\{\sum_{x,y}\bar{\mathbf{C}}_5^{\times}(x,y,\Delta x,\Delta y)\middle|[\Delta x,\Delta y]\in\{(0,1),(1,1),(1,-1),(0,8),(8,8)\}\right\}$,
2. $\mathcal{I}_{\mathrm{f}}^{\star}=\left\{\sum_{x,y}\hat{\mathbf{C}}_5^{\star}(x,y,\Delta x,\Delta y)\middle|[\Delta x,\Delta y]\in\{(0,1),(1,0),(1,1),(1,-1)\}\right\},\star\in\{\rightarrow,\downarrow,\searrow,\rightrightarrows,\downdownarrows\}$,
3. $\mathcal{I}_{\mathrm{s}}^{\star}=\left\{\sum_{x,y}\hat{\mathbf{C}}_5^{\star}(x,y,\Delta x,\Delta y)\middle|[\Delta x,\Delta y]\in\{(0,8),(8,0),(8,8),(8,-8)\}\right\},\star\in\{\rightarrow,\downarrow,\searrow,\rightrightarrows,\downdownarrows\}$,

对于块内对，上述定义的总和总是对所有 DCT 模式(x,y)∈{0，···,7}^{2}， 所以(x, y)和(x+Δx y+Δy)位于相同的 8×8 块。当索引最终出现在 DCT 数组之外时，块间矩阵也会受到类似的约束。所有定义中都省略了直流模式。

子模型 $\mathcal{I}^{\times}$讨论了空间(块间)和频率(块内)的依赖关系，可以看作是 Liu[14]对 absNJ1 和 absNJ2 特性集的扩展。基于差分的子模型与[1]中提出的马尔科夫特征相似，作者还利用了 DCT 系数绝对值之间的块间和块内差异。为了获得类似的维数，从差异中计算出的共现被分成两个不同的集合，分别捕获频率($\mathcal{I}_{\mathrm{f}}^{\star}$)和空间($\mathcal{I}_{\mathrm{s}}^{\star}$)相关性。

将具体的 DCT 模式子模型与完整的子模型相结合，形成了 JPEG 域丰富模型。它的总维数是 11255。为了提高性能，我们采用了二倍量纲 22,510 的笛卡尔坐标校准[8]。整个 JRM 的结构如图 1 所示。

3. COMPARISON TO PRIOR ART

To demonstrate the power of the proposed JPEG rich model, we steganalyze six modern steganographic methods. We use the ensemble classifier [9, 10] for all experiments as it enables fast training in high-dimensional feature spaces and its performance on low-dimensional feature sets is comparable to the much more complex SVMs [10]. The classifier is constructed from base learners implemented as Fisher Linear Discriminants on random subspaces of the feature space. The number of subspaces and their dimensionality were determined automatically using the algorithms described in [10]. A publicly available implementation of the ensemble classifier can be downloaded from http://dde.binghamton.edu/download/ensemble.
为了证明所提出的 JPEG 富模型的威力，我们分析了六种现代隐写术方法。我们在所有实验中都使用集成分类器[9,10]，因为它可以在高维特征空间中进行快速训练，并且在低维特征集上的性能可以与更复杂的 SVMs[10]相媲美。该分类器由基于特征空间的随机子空间上的 Fisher 线性鉴别器实现的基学习器构成。使用[10]中描述的算法自动确定子空间的数量及其维数。可以从 http://dde.binghamton.edu/download/ensemble.上面下载集成分类器的公开实现

3.1 Tested steganographic methods （隐写术检测方法）

The tested steganographic methods are: nsF5, MBS, YASS, MME, BCH, and BCHopt. The nsF5 algorithm [5] is an improved version of the popular F5 [24]. For experiments, we used a simulator of nsF5, available at http://dde.binghamton.edu/download/nsf5simulator/, that makes the embedding changes as if an optimal binary matrix coding scheme was used. We note that a near-optimal practical implementation can be achieved using syndrome-trellis codes [2].

测试的隐写方法有：nsF5, MBS, YASS, MME, BCH, BCHopt。nsF5 算法[5]是流行的 F5[24]的改进版本。在实验中，我们使用了 nsF5 的模拟器，可以在 http://dde.binghamton.edu/download/nsf5simulator/上找到，这使得嵌入的变化就像使用了最佳的二进制矩阵编码方案一样。我们注意到，使用综合症格架代码[2]可以实现接近最优的实际实现。

MBS is a model-based steganography due to Sallee [19]. The implementation we used is available at http://www.philsallee.com/mbsteg. Both nsF5 and MBS start directly with the JPEG image to be modified and thus do not utilize any side information.
由于 Sallee [19]， MBS 是一种基于模型的隐写术。我们使用的实现可在 http:/ /www.philsallee.com/mbsteg。nsF5 和 MBS 都直接从要修改的 JPEG 图像开始，因此不利用任何边信息。

YASS, which hides data robustly in a transform domain, was introduced in [23] and later improved in [20,21]. Even though it is easily detectable today [11, 13, 14], it played an important role to clarify the real purpose of the process of feature calibration [8]. We test five different settings of YASS, numbered 3, 8, 10, 11, and 12 in [11], as these were reported to be the most secure. YASS performs only full embedding and thus the reported payloads are averages over all images in the CAMERA database to be described next.

YASS 在一个变换域中稳健地隐藏数据，它是在[23]中引入的，后来在[20,21]中进行了改进。虽然现在很容

易检测到[11,13,14]，但它对阐明特征校准[8]过程的真正目的起到了重要作用。我们测试了五种不同的 YASS 设置，分别是[11]中的 3、8、10、11 和 12，因为这些设置据说？？？是最安全的。YASS 只执行完全嵌入，因此所报告的有效载荷是下面将要描述的摄像机数据库中所有图像的平均值。

MME [6] utilizes side information at the sender in terms of the uncompressed image and employs matrix embedding to minimize an appropriately defined distortion function. We tested its Java implementation that uses a Java JPEG encoder for image compression. Therefore, in order to steganalyze solely the impact of MME embedding, we need to create cover images using the same compressor to avoid artificially increasing the detection reliability by also detecting traces of a different JPEG compressor [7].

MME[6]利用发送方未压缩图像的侧信息，并使用矩阵嵌入来最小化适当定义的失真函数。我们测试了它使用 Java JPEG 编码器进行图像压缩的 Java 实现。因此，为了仅隐写分析 MME 嵌入的影响，我们需要使用相同的压缩器创建覆盖图像，以避免通过检测不同 JPEG 压缩器[7]的踪迹来人为地提高检测的可靠性。

BCH and BCHopt [18] are side-informed algorithms that employ BCH codes to minimize the embedding distortion in the DCT domain defined using the knowledge of non-rounded DCT coefficients. BCHopt is an improved version of BCH that contains a heuristic optimization and also hides message bits into zeros. According to the experiments in [18], BCHopt is currently the most secure practical JPEG steganographic scheme. To the best of our knowledge, it has not been steganalyzed elsewhere.

BCH 和 BCHopt[18]是采用 BCH 编码的侧知算法，利用非四舍五入 DCT 系数的知识来最小化 DCT 域中的嵌入失真。BCHopt 是 BCH 的改进版本，它包含一个启发式优化，并且将消息位隐藏为零。根据[18]的实验，BCHopt 是目前最安全实用的 JPEG 隐写方案。据我们所知，还没有人在其他地方对此进行过分析。

3.2 Performance evaluation

The image source on which all experiments were carried out is the CAMERA database containing 6,500 JPEG images originally acquired in their RAW format taken by 22 digital cameras, resized so that the smaller size is 512 pixels with aspect ratio preserved, and converted to grayscale. The cover images for MME were created using a Java JPEG encoder. For the rest, we used Matlab's function imwrite. The JPEG quality factor was fixed to 75 in both cases.

所有实验进行的图像来源是包含 6500 张原始格式的 JPEG 图像的相机数据库，由 22 台数码相机拍摄，调整大小使较小的尺寸为 512 像素，保留高宽比，并转换为灰度。MME 的封面图像是使用 Java JPEG 编码器创建的。剩下的部分，我们使用 Matlab 的 imwrite 函数。在这两种情况下，JPEG 质量因子都被固定为 75。

For every steganographic method, we created stego images using a range of different payload sizes expressed in terms of bits per nonzero AC DCT coefficient (bpac), and trained a separate classifier to detect each of them. Before classification, all cover-stego pairs were divided into two halves for training and testing, respectively. We define the minimal total error P_{E} under equal priors achieved on the testing set as

对于每一种隐写术方法，我们使用以每个非零交流 DCT 系数(bpac)位表示的不同有效载荷大小范围来创建隐写术图像，并训练一个单独的分类器来检测它们。在分类之前，所有的 cover-stego 对被分成两半，分别进行训练和测试。我们定义测试集上相同先验条件下的最小总误差 PE 为

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{P_{\mathrm{FA}} + P_{\mathrm{MD}}(P_{\mathrm{FA}})}{2}, \qquad (14)$$

where P_{FA} is the false alarm rate （误报率） and P_{MD} is the missed detection rate（漏检率）. The performance is evaluated using the median value of P_{E} over ten random 50/50 splits of the database and will be denoted as $\bar{P}_{\mathrm{E}}$. 性能是用 P_{E} 的中值除以数据库的 10 次随机 50/50 分割来评估的，并将其表示为 P_{E}。

We compare the steganalysis performance of the following feature spaces (models); the numbers in brackets denote their dimensionality:

比较了以下特征空间(模型)的隐写分析性能；括号内的数字表示它们的维数：

- CHEN (486) = Markov features utilizing both intra- and inter-block dependencies [1], 利用块内和块间相关性的马尔科夫特性
- CC-CHEN (972) = CHEN features improved by Cartesian calibration [8], CHEN 特征通过笛卡尔校正得到改善
- LIU (216) = the union of *diff-absNJ-ratio* and *ref-diff-absNJ* features published in [14], 关于 diffabsNJ -ratio 与 ref- diffabsNJ 特征的联合发表于
- CC-PEV (548) = Cartesian-calibrated PEV feature set [17], 笛卡尔校准的 PEV 特征集

- CDF (1,234) = CC-PEV features expanded by SPAM features [16] extracted from spatial domain, 利用空间域提取的垃圾邮件特征[16]扩展 CC-PEV 特征，
- CC-C300 (48,600) = the high-dimensional feature space proposed in [9], [9]中提出的高维特征空间
- CF^* (7,850) = compact rich model for DCT domain proposed in [10], [10]中提出的 DCT 域的紧致富模型，
- JRM (11,255) = the rich model proposed in this paper, without calibration, 本文提出的 rich 模型无需标定，
- CC-JRM (22,510) = Cartesian-calibrated JRM, 笛卡尔-校准 JRM
- J+SRM (35,263) = the union of CC-JRM and the Spatial-domain Rich Model (SRM) proposed in [4] (all 39 submodels of SRM were taken with a fixed quantization $q = 1c$, see [4] for more details). CC-JRM 与[4]中提出的空间域丰富模型(SRM 的 39 个子模型均采用固定量化 q = 1c，详见[4])的结合。

Resulting errors $\bar{P}_E$ are reported in Table 1. The proposed CC-JRM delivers the best performance among all feature sets that are extracted directly from the DCT domain, across all tested steganographic methods and all payloads. Adding the spatial-domain rich model [4] further improves the performance and delivers the overall best results.

因此产生的误差 P¯E 见表 1。所提出的 CC-JRM 在所有直接从 DCT 域提取的特征集中，在所有经过测试的隐写方法和所有有效载荷下，提供了最好的性能。添加空间域丰富的模型[4]进一步提高了性能，并提供了整体的最佳结果。

Algorithm	Payload (bpac)	CHEN (486)	CC-CHEN (972)	LIU (216)	CC-PEV (548)	CDF (1,234)	CC-C300 (48,600)	CF* (7,850)	JRM (11,255)	CC-JRM (22,510)	J+SRM (35,263)
nsF5	0.050	0.4153	0.3816	0.3377	0.3690	0.3594	0.3722	0.3377	0.3407	0.3298	0.3146
	0.100	0.3097	0.2470	0.1732	0.2239	0.2020	0.2207	0.1737	0.1782	0.1616	0.1375
	0.150	0.2094	0.1393	0.0706	0.1171	0.0906	0.1127	0.0720	0.0793	0.0663	0.0468
	0.200	0.1345	0.0708	0.0273	0.0549	0.0360	0.0486	0.0273	0.0338	0.0255	0.0150
MBS	0.010	0.4070	0.3962	0.3826	0.3876	0.3786	0.4038	0.3710	0.3478	0.3414	0.3260
	0.020	0.3178	0.2962	0.2780	0.2827	0.2684	0.3120	0.2560	0.2156	0.2122	0.1832
	0.030	0.2395	0.2100	0.1925	0.1965	0.1795	0.2241	0.1684	0.1266	0.1195	0.0983
	0.040	0.1770	0.1437	0.1288	0.1298	0.1135	0.1594	0.1087	0.0751	0.0670	0.0494
	0.050	0.1243	0.0946	0.0812	0.0833	0.0704	0.1176	0.0684	0.0427	0.0373	0.0282
YASS (12)	0.077	0.2009	0.1825	0.2324	0.2279	0.1268	0.0930	0.0532	0.0324	0.0303	0.0173
YASS (11)	0.114	0.1989	0.1585	0.2118	0.1573	0.0718	0.0701	0.0437	0.0349	0.0227	0.0111
YASS (8)	0.138	0.2520	0.1911	0.1886	0.1827	0.0742	0.0500	0.0271	0.0287	0.0178	0.0104
YASS (10)	0.159	0.2334	0.1476	0.1793	0.1341	0.0507	0.0370	0.0164	0.0210	0.0103	0.0054
YASS (3)	0.187	0.1277	0.0876	0.1301	0.0723	0.0224	0.0350	0.0146	0.0165	0.0081	0.0045
MME	0.050	0.4678	0.4546	0.4479	0.4492	0.4340	0.4427	0.4443	0.4424	0.4307	0.4194
	0.100	0.3001	0.2611	0.2574	0.2613	0.2501	0.3026	0.2466	0.2286	0.2091	0.1891
	0.150	0.2165	0.1735	0.1677	0.1721	0.1586	0.2299	0.1608	0.1404	0.1221	0.1027
	0.200	0.0217	0.0104	0.0127	0.0127	0.0124	0.0726	0.0153	0.0112	0.0080	0.0059
BCH	0.100	0.4599	0.4496	0.4448	0.4426	0.4390	0.4497	0.4290	0.4305	0.4229	0.4060
	0.200	0.3594	0.3124	0.3087	0.2974	0.2752	0.2958	0.2629	0.2707	0.2369	0.1946
	0.300	0.1383	0.0889	0.0862	0.0779	0.0697	0.0912	0.0663	0.0715	0.0536	0.0390
BCHopt	0.100	0.4726	0.4683	0.4558	0.4618	0.4595	0.4684	0.4550	0.4515	0.4480	0.4306
	0.200	0.4032	0.3712	0.3583	0.3548	0.3368	0.3517	0.3265	0.3253	0.3030	0.2582
	0.300	0.2400	0.1711	0.1719	0.1605	0.1356	0.1681	0.1289	0.1389	0.1102	0.0830

Table 1. Median testing error $\bar{P}_E$ for six JPEG steganographic methods using different models. For easier navigation, the gray-level of the background in each row corresponds to the performance of individual feature sets: darker ⇒ better performance (lower error rate).

使用不同模型的六种 JPEG 隐写方法的中间测试误差 P¯E。为了便于导航，每一行的背景灰度值对应于各特征集的性能：格深更好性能(较低的错误率)。

3.3 Discussion（讨论）

Table 1 provides an important insight into the feature-building process and points to the following general guidelines for design of feature spaces for steganalysis:

表 1 对功能构建过程提供了重要的见解，并指出了以下用于隐写分析的功能空间设计的一般指导原则:

- *High dimension is not sufficient for good performance*. This is clearly demonstrated by the rather poor performance of the 48,600-dimensional CC-C300 feature set, often outperformed by the significantly lower-dimensional sets LIU, CC-PEV, and CC-CHEN. The failure of CC-C300 could be attributed to its lack of diversity (all co-occurrences are of the same type) and missing symmetrization, which make the model less robust and unnecessarily high-dimensional.

  高维度对良好的性能来说是不够的。这一点从 48,600 维的 CC-C300 特性集的相当糟糕的性能中得到了清楚的证明，它们的性能常常被更低的维度集 LIU、CC-PEV 和 CC-CHEN 所超越。CC-C300 的失败可能是由于其缺乏多样性(所有的共现都是同一类型)和缺少对称，这使得模型不那么健壮和不必要的高维。
- *Calibration helps*. The positive effect of calibration has been demonstrated many times in the past, and here we confirm it by comparing the columns CHEN → CC-CHEN and JRM → CC-JRM. Notice that even for the high-dimensional JRM, the improvement may be substantial: 0.2707 → 0.2369 for BCH at 0.2 bpac and 0.1404 → 0.1221 for MME at 0.15 bpac. Moreover, researching alternative ways of calibration may bring additional improvements to feature-based steganalysis. This is indicated by a relatively good performance of the LIU feature set (compared to other low-dimensional sets), which utilizes two novel calibration principles: strenghtening the reference statistics by averaging over 63 different image croppings and calibrating by the *ratio* between original and reference features [14].

  校准有帮助。校准的积极作用在过去已经被证明了很多次，在这里我们通过比较 CHEN→CC-CHEN 和

JRM→CC-JRM 两栏来确认它。请注意，即使对于高维 JRM，改进也可能是显著的：BCH 在 0.2 bpac 时为 0.2707→0.2369,MME 在 0.15 bpac 时为 0.1404→0.1221。此外，还研究了多种标定方法可能会给基于特征的隐写分析带来额外的改进。与其他低维集相比，LIU 特征集的性能相对较好，这表明了这一点。LIU 特征集利用了两种新的校准原则：通过平均 63 种不同的图像裁剪来加强参考统计，并通过原始和参考特征之间的比率[14]校准。

- *Steganalysis benefits from cross-domain models*. By combining CC-PEV with the spatial-domain SPAM features (CDF), sizeable improvement over CC-PEV is apparent across all steganographic methods. The

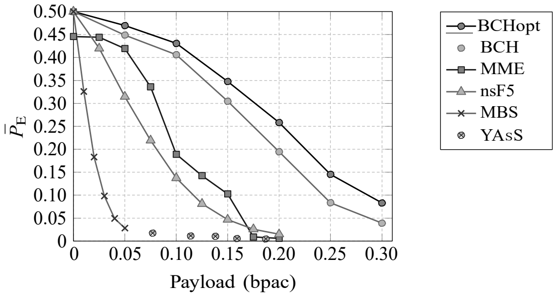


Figure 2. Testing error $\bar{P}_E$ using the J+SRM feature space (dimension 35,263).

benefit of the multiple-domain approach is also clear from the last column – the union of the CC-JRM and the spatial domain rich model proposed in [4] further markedly improves the performance of CC-JRM and yields the lowest achieved error rates in all cases.
隐写分析得益于跨域模型。通过将 CC-PEV 与空间域垃圾邮件特性(CDF)相结合，所有隐写术方法都明显优于 CC-PEV。多域方法的好处在最后一列中也很明显——CC-JRM 和[4]中提出的空间域丰富模型的结合进一步显著提高了 CC-JRM 的性能，并在所有情况下获得最低的出错率。

- *Future steganalysis will likely be driven by diverse and compact rich models*. The systematically constructed JPEG rich models CF^* and JRM/CC-JRM consistently outperform all low-dimensional sets. The superior performance of CC-JRM over CF^* is due to additional symmetrization, further diversification by co-occurrences of *differences*, and by its new integral components.
•未来的隐写分析可能会由多样化和紧凑的富模型驱动。系统构建的 JPEG 丰富模型 CF 被显示出来，而 JRM/CC-JRM 始终优于所有低维集。相对于 CF*来说，CC-JRM 的优越性能是由于额外的对称化和 co-的进一步多样化造成的出现的差异，并由其新的组成部分。

3.4 Comparison of steganographic methods

In Figure 2, we compare the performance of all tested stego schemes using the J+SRM feature set. We confirm that BCHopt is the most secure steganographic method, and its heuristic optimization brings improvement over BCH.

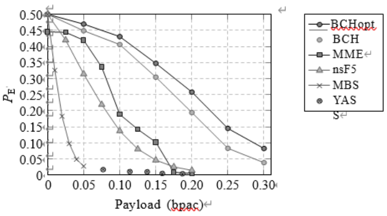
MBS and YASS are by far the least secure algorithms. The failure of YASS, already reported in [11, 14], suggests that embedding robustly in a different domain may not be the best approach for passive warden steganography as the robustness unavoidably yields significant and thus easily detectable distortion.

The nsF5 algorithm, which does not utilize any side information, is clearly outperformed by all schemes that utilize the knowledge of the uncompressed cover: MME, BCH, and BCHopt. The effect of this type of side information at the sender on steganographic security is, however, not well understood today. In particular, it is not clear how to utilize it in the best possible manner.

在图 2 中，我们使用 J+SRM 特征集比较了所有经过测试的 stego 方案的性能。我们确认 BCHopt 是最安全的隐写方法，其启发式优化带来了对 BCH 的改进。

到目前为止，MBS 和 YASS 是最不安全的算法。YASS 的失败(已经在[11,14]中报道过)表明，将粗鲁性嵌入到不同的领域可能不是被动的典狱长隐写术的最佳方法，因为粗鲁性不可避免地会产生显著的、因此很容易检测到的失真。

nsF5 算法，它没有利用任何边信息，显然是优于所有方案，利用知识的未压缩封面:MME，BCH，和 BCHopt。然而，发送方的这类侧信息对隐写术安全性的影响目前还没有得到很好的理解。特别是，不清楚如何以最佳方式利用它。



Let us conclude this section by commenting on two security artifacts of MME. First, we can see significant jumps in P_E around payloads 0.09 and 0.16 bpac, which are due to suboptimal Hamming codes as already reported in [11]. This could be easily remedied by using more sophisticated coding schemes [2]. Second, note that the error at zero payload is $P_E \approx 0.45$ rather than random guessing. This is caused by the embedded message header whose size is independent of the message length and which is present in every stego image. We found that in case of MME, this message header is always embedded in the top left corner of the image, in many cases the area of sky, and thus creates statistically detectable traces. Even though this implementation flaw could be easily fixed, it illustrates that even the smallest implementation detail needs to be handled with caution when designing a practical steganographic scheme.

首先，我们可以看到 P¯E 在有效载荷 0.09 和 0.16 bpac 附近有显著的跳跃，这是由于汉明码不够优化造成的在[11]。通过使用更复杂的编码方案[2]可以很容易地纠正这一点。其次，请注意，零有效载荷下的误差是 P¯E≈0.45，而不是随机猜测。这是由嵌入式引起的消息头，其大小与消息长度无关，并且存在于每个 stego 图像中。我们发现，在 MME 的情况下，这个消息头总是嵌入在图像的左上角，在许多情况下是天空的区域，从而创建统计上可检测到的踪迹。尽管这个实现缺陷很容易修复，但它说明在设计实用的隐写方案时，即使是最小的实现细节也需要谨慎处理。

4. INVESTIGATIVE EXPERIMENTS

The purpose of this section is to study the contribution of the individual components of the CC-JRM to the overall performance. We also address the problem of finding a small subset of CC-JRM responsible for most of the detection accuracy for a fixed stego source. Our experiments are restricted only to selected steganographic methods and payloads. The notation follows Figure 1.

本节的目的是研究 CC-JRM 的各个组件对总体性能的贡献。我们还解决了寻找 CC-JRM 的一个小子集的问题，该子集负责对固定 stego 源的大部分检测精度。我们的实验仅限于选定的隐写术方法和有效载荷。符号如下图 1 所示。

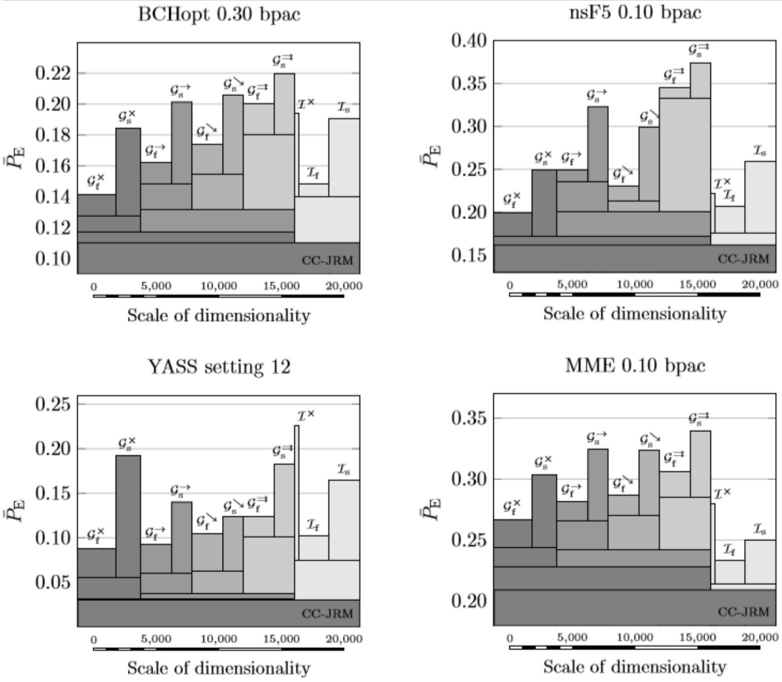


Figure 3. Systematic merging of the CC-JRM submodels and the progress of the testing error $\bar{P}_\mathrm{E}$. See Section 4.1 for explanation of the graphs and their interpretation.

图 3。CC-JRM 子模型的系统合并和测试误差 P¯E 的进展。图的解释和解释见第 4.1 节。

$$P_\mathrm{E} = \min_{P_\mathrm{FA}} \frac{P_\mathrm{FA} + P_\mathrm{MD}(P_\mathrm{FA})}{2},$$

4.1 Systematic merging of submodels

In the first experiment, we consider the following disjoint and qualitatively different subsets of the CC-JRM: $\mathcal{G}_\mathrm{f}^\times, \mathcal{G}_\mathrm{s}^\times, \mathcal{G}_\mathrm{f}^\rightarrow, \mathcal{G}_\mathrm{s}^\rightarrow, \mathcal{G}_\mathrm{f}^\searrow, \mathcal{G}_\mathrm{s}^\searrow, \mathcal{G}_\mathrm{f}^\rightrightarrows, \mathcal{G}_\mathrm{s}^\rightrightarrows, \mathcal{I}^\times, \mathcal{I}_\mathrm{f}, \mathcal{I}_\mathrm{s},$, and use them for steganalysis separately. Afterwards,we gradually and systematically merge them together, following the logic of Figure 1, until all of them are merged into the CC-JRM. All considered feature sets are Cartesian calibrated, yielding double the dimensionalities shown in Figure 1. The experiment was performed on the following steganographic schemes: BCHopt 0.30 bpac, nsF5 0.10 bpac, YASS setting 12, and MME 0.10 bpac. The training procedure was identical to the one used in the experiments of Section 3: training on a randomly selected half of the CAMERA database, testing on the other half. The obtained performance is reported in Figure 3 in terms of $\bar{\mathrm{P}}_\mathrm{E}$.

在第一个实验中，我们考虑以下 CC-JRM 的不同子集的不相交和性质: $\mathcal{G}_\mathrm{f}^\times, \mathcal{G}_\mathrm{s}^\times, \mathcal{G}_\mathrm{f}^\rightarrow, \mathcal{G}_\mathrm{s}^\rightarrow, \mathcal{G}_\mathrm{f}^\searrow, \mathcal{G}_\mathrm{s}^\searrow, \mathcal{G}_\mathrm{f}^\rightrightarrows, \mathcal{G}_\mathrm{s}^\rightrightarrows, \mathcal{I}^\times, \mathcal{I}_\mathrm{f}, \mathcal{I}_\mathrm{s},$，并分别使用它们进行隐写分析。然后，按照图 1 的逻辑，我们逐步地、系统地将它们合并在一起，直到所有这些都合并到 CC-JRM 中。所有考虑的特征集都经过了笛卡尔校准，产生了图 1 所示的两倍的维数。实验采用以下隐写方案:

BCHopt 0.30 bpac、nsF5 0.10 bpac、YASS set 12、MME 0.10 bpac。训练过程与第 3 部分的实验相同：在相机数据库中随机选择一半进行训练，在另一半进行测试。获得的性能在图 3 中以 P¯E 表示。

In Figure 3, every submodel is represented by a bar whose height is the $\bar{\mathsf{P}}_{\mathrm{E}}$. Conveniently, the width of each bar is proportional to the dimensionality of the corresponding submodel, allowing thus a continuous perception of the feature space sizes. For example, the rather thin bar of $\mathcal{I}^{\times}$ can be immediately perceived as more than five times smaller than the neighboring I_{f}. Intuitively, the union of several submodels is represented by an overlapping bar whose width is equal to the sum of its components. The overlapping bars do not interfere with the performance of their submodels because merging always decreases the error. For example, see the performance of submodels $\mathsf{G}_{\mathrm{f}}^{\Rightarrow}$ and $\mathsf{G}_{\mathrm{s}}^{\Rightarrow}$ in the top left graph (BCHopt). Their individual errors P_{E} are 0.20 and 0.22, respectively, and their union (denoted $\mathsf{G}^{\Rightarrow}$ in Figure 3) yields error 0.18, thence the corresponding height of the lower, wider bar. After adding additional submodels $\mathsf{G}^{\rightarrow}$,$\mathsf{G}^{\rightarrow}$,G�,G�, the error can be seen to drop further to roughly 0.13. The final performance of the CC-JRM is always represented by the lowest bar spanning the whole width of the graph. Finally, the readability is further improved by using different shades of gray for different types of submodels.

在图 3 中，每个子模型都用一个高度为 P¯E 的条表示。每个条的宽度与相应子模型的维数成正比，从而允许连续的感知特征空间的大小。例如，$\mathcal{I}^{\times}$的较细的条可以立即被感知为大于比相邻的 If 小五倍。直观地说，几个子模型的并集由一个宽度等于其各分量之和的重叠条。重叠的条不干涉它们的子模型的性能，因为合并总是减少错误。例如，在左上角的图(BCHopt)中可以看到$\mathcal{G}_{\mathrm{f}}^{\rightrightarrows}$和$\mathcal{G}_{\mathrm{s}}^{\rightrightarrows}$的子模型的性能。他们的单个误差 P¯E 为 0.20 和 0.22 分别，在图 3 中，$\mathcal{G}^{\rightrightarrows}$为它们的并集，因此产生 0.18 的误差，从而得出相应的高度较低，较宽的。在添加额外的子模型$\mathcal{G}_{\mathrm{f}}^{\rightarrow},\mathcal{G}_{\mathrm{s}}^{\rightarrow},\mathcal{G}_{\mathrm{f}}^{\searrow},\mathcal{G}_{\mathrm{s}}^{\searrow}$,之后，误差将进一步下降到大约 0.13。CC-JRM 的最终性能总是由跨越整个图形宽度的最低条表示。最后，通过对不同类型的子模型使用不同的灰度，进一步提高了模型的可读性。

Figure 3 reveals interesting information about the types of features that are effective for attacking various steganographic algorithms. The four selected steganographic methods represent very different embedding paradigms, which is why the individual submodels contribute differently to the detection. The contribution of the integral features $\mathcal{I}=\{\mathcal{I}^{\times},\mathcal{I}_{\mathrm{f}},\mathcal{I}_{\mathrm{s}}\}$,for example, seems to be rather negligible for YASS because steganalysis *without* I delivers basically the same performance. For the other three algorithms, however, integral features noticeably improve the performance. This is most apparent for MME where the integral features I perform better than the rest of the features together despite their significantly lower dimensionality. As another example, compare the individual performance of the DCT-mode specific features extracted directly from absolute， values of DCT coefficients $\mathcal{G}^{\times}=\{\mathcal{G}_{\mathrm{f}}^{\times},\mathcal{G}_{\mathrm{s}}^{\times}\}$,with the DCT-mode specific features extracted from the differences, $\mathcal{G}_{\mathrm{diff}}=\{\mathcal{G}_{\mathrm{f}}^{\rightarrow},\mathcal{G}_{\mathrm{s}}^{\rightarrow},\mathcal{G}_{\mathrm{f}}^{\searrow},\mathcal{G}_{\mathrm{s}}^{\searrow},\mathcal{G}_{\mathrm{f}}^{\rightrightarrows},\mathcal{G}_{\mathrm{s}}^{\rightrightarrows}\}$. While for nsF5, Gdiff does not improve the performance of $\mathcal{G}^{\times}$ much, bothseem to be important for the other three algorithms and especially for YASS.

图 3 显示了关于攻击各种隐写算法的有效特征类型的有趣信息。所选的四种隐写术方法代表了非常不同的嵌入范式，这就是为什么各个子模型对检测的贡献不同。$\mathcal{I}=\{\mathcal{I}^{\times},\mathcal{I}_{\mathrm{f}},\mathcal{I}_{\mathrm{s}}\}$,的贡献例如，整体特征对于 YASS 来说似乎是微不足道的，因为隐写分析没有 I，性能基本相同。而对于其他三种算法，则是积分特征显著提高性能。这是最明显的对 MME 的整合功能，尽管它们的维数很低，但比其他功能加在一起更好。另外一个例子,比较 DCT-mode 特定的个人绩效直接从绝对的特征提取,DCT 系数的值,与 DCT-mode 从差异中提取特定的功能,而对于 nsF5，Gdiff（$\mathcal{G}_{\mathrm{diff}}=\{\mathcal{G}_{\mathrm{f}}^{\rightarrow},\mathcal{G}_{\mathrm{s}}^{\rightarrow},\mathcal{G}_{\mathrm{f}}^{\searrow},\mathcal{G}_{\mathrm{s}}^{\searrow},\mathcal{G}_{\mathrm{f}}^{\rightrightarrows},\mathcal{G}_{\mathrm{s}}^{\rightrightarrows}\}$.）并不能提高性能的多,两者看似对其他三个算法很重要,尤其是 YASS。

We conclude that there is no subset of CC-JRM that is universally responsible for majority of detection accuracy across different steganographic schemes. The power of CC-JRM is in the *union* of its systematically built components, carefully designed to capture different types of statistical dependencies.

我们的结论是，在不同的隐写方案中，CC-JRM 没有一个子集是普遍负责大多数检测精度的。CC-JRM 的**强大之处在于其系统构建的组件的联合**，精心设计以捕获不同类型的统计相关性。

4.2 Forward feature selection

Despite its high dimension (22,510), ensemble classifiers make the training in the CC-JRM feature space computationally feasible. In fact, the bottleneck of steganalysis now becomes the feature extraction rather than the actual training of the classifier. To give the reader a better idea, we measured the running time needed for steganalysis of nsF5 at 0.10 bpac using the CC-JRM. The extraction of features from 6,500 images took roughly 18 hours, while, on the same machine,[†] the classifier training took on average 5 minutes. From the practical point of view, the *testing* time may be an important factor – after the classifier is trained, the time needed to make decisions should be minimized.

Although projecting the CC-JRM feature vector of the image under inspection into eigen-directions of individual FLDs of the ensemble classifier consists of a series of fast matrix multiplications, the extraction of the 22,510 complex features is quite costly. Therefore, one may want to consider investing more time into the training procedure, and perform a supervised feature selection in order to reduce the number of features needed to be extracted during testing, while keeping satisfactory performance. Note that we are interested specifically in feature selection rather than general dimensionality-reduction as the goal is to minimize the number of features needed.

尽管集成分类器具有较高的维数(22,510)（经过笛卡尔校准），但它使 CC-JRM 特征空间模型的训练在理论上是可行的。事实上，隐写分析的瓶颈现在变成了**特征提取**，而不是分类器的实际训练。为了让读者更好地理解，我们测量了所需的运行时间使用 CC-JRM 对 0.10 bpac 时的 nsF5 进行隐写分析。从 6500 张图像中提取特征大约 18 个小时，而在相同的机器上，分类器训练平均花费 5 分钟。从实用的角度来看，测试时间可能是一个重要的因素——在训练分类器之后，应该尽量减少决策所需的时间。

虽然将被检图像的 CC-JRM 特征向量投影到集成分类器的各个 FLDs 的特征方向上需要一系列快速的矩阵乘法，但是提取 22510 个复杂特征的代价非常高。因此，可以考虑在训练过程中投入更多的时间，进行监督特征选择，以减少测试中需要提取的特征数量，同时保持令人满意的性能。请注意，我们特别感兴趣的是**特性选择**，而不是一般的降维，因为目标是最小化必需特性的数量。

Unfortunately, as shown in the previous investigative experiment in Figure 3, there is no compact subset of CC-JRM that would be universally effective against different types of embedding modifications. However, if we *fix* the steganographic channel, the problem becomes feasible. To demonstrate the feasibility of this direction, we performed a rather simple forward feature selection procedure applied to submodels (it was called the ITERATIVE-BEST strategy in [4]). It starts with all $N = 2 \times 51 = 102$ submodels of the CC-JRM.[‡] Once $k \geq 0$ submodels are selected, add the one that leads to the biggest drop in the OOB error estimate when all $k+1$ submodels are used as a feature space. We use the out-of-bag (OOB) error estimation calculated from the training set [10] because no information about the testing images can be utilized. The procedure finishes after a pre-defined number of iterations or once the OOB values reach satisfactory values.

不幸的是，正如图 3 中先前的研究实验所示，CC-JRM 没有一个紧凑的子集可以普遍有效地抵抗不同类型的嵌入修改。然而，如果我们修正隐写通道，问题就变得可行了。为了证明这个方向的可行性，我们执行了一个应用于子模型的相当简单的正向特征选择过程[4]中的迭代最佳策略)。它从 CC-JRM 的所有 N=2×51=102 个子模型开始。一旦选择 k≥0 的子模型，将导致 OOB（袋外数据 out of bag）误差估计下降最大的子模型相加将 k + 1 个子模型作为特征空间。我们使用了 out-of-bag (OOB)错误估计训练集[10]，因为没有关于测试图像的信息可以利用。程序完成后预先定义的迭代次数或一次 OOB 值达到令人满意的值。

批注 [.皓4]: 建议利用 OOB error 估计作为**泛化误差估计**的一个组成部分,并且 Breiman 在论文中给出了经验性实例表明袋外数据误差估计与同训练集一样大小的测试集得到的精度一样，这样也就表明袋外数据(oob)误差估计是一种可以取代测试集的误差估计方法。

对于已经生成的随机森林，用袋外数据测试其性能，假设**袋外数据总数为 O**，用这 O 个袋外数据作为输入，带进之前已经生成的随机森林分类器,分类器会给出 O 个数据相应的分类，因为这 O 条数据的类型是已知的，则用正确的分类与随机森林分类器的结果进行比较，**统计随机森林分类器分类错误的数目，设为 X**，则袋外数据误差大小=X/O；这已经经过证明是无偏估计的，所以在随机森林算法中不需要再进行交叉验证或者单独的测试集来获取测试集误差的无偏估计。

The ITERATIVE-BEST strategy greedily minimizes the OOB error at every iteration and takes the mutual dependencies among individual submodels into account. The ensemble classifier is used as a black box providing classification feedback. Such methods are known as wrappers [12].

We performed the ITERATIVE-BEST feature selection strategy on BCHopt 0.30 bpac, nsF5 0.10 bpac, YASS setting 12, and MME 0.10. The results are shown in Figure 4. The individual graphs show the progress of OOB error estimates for $k \leq 10$ as well as the list of the selected submodels. We follow the notation of the submodels

[†] Dell PowerEdge R710 with 12 cores and 48GB RAM when executed as a single process.

[‡] We treat the submodels and their reference submodels coming from Cartesian calibration separately.

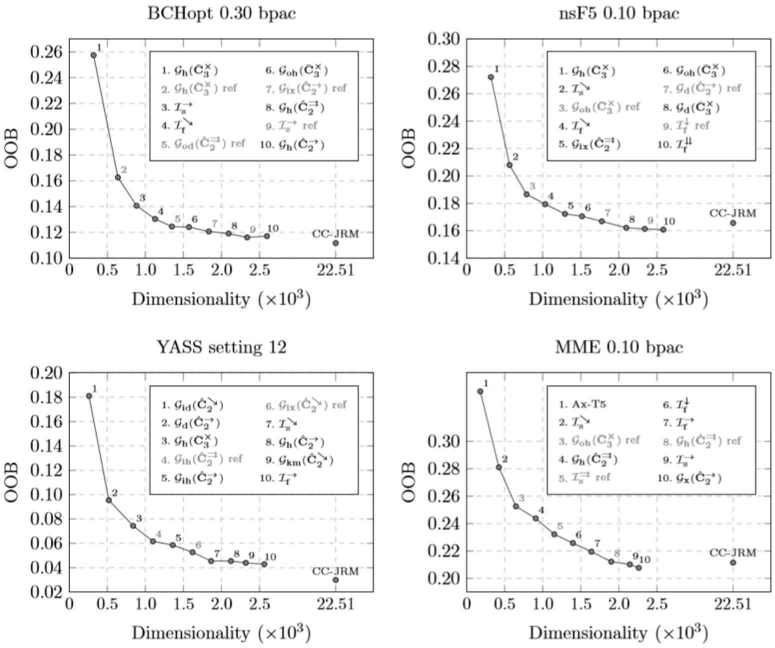


Figure 4. The result of the ITERATIVE-BEST feature selection strategy applied to four qualitatively different steganographic methods. The reported values are out-of-bag (OOB) error estimates calculated on the training set (half of the CAMERA database). For reference, we include the OOB error of the full CC-JRM feature space.
迭代最佳特征选择策略的结果应用于四种不同性质的隐写方法。报告的值是根据训练集(摄像机数据库的一半)计算的 out-of-bag (OOB)错误估计。作为参考，我们包含了整个 CC-JRM 特征空间的 OOB 错误。

批注 [.皓5]: 图 4：维度越高，OOB 袋外错误越小

introduced in Figure 1 and distinguish the reference-version of the submodels by gray shade and adding the suffix “ref.” For comparison, we also include the OOB-performance when the entire CC-JRM is used.
迭代最佳策略贪婪地将每次迭代的 OOB 错误最小化，并考虑到各个子模型之间的相互依赖关系。集成分类器作为一个黑盒提供分类反馈。这种方法称为包装器[12]。我们在 BCHopt 0.30bpac、nsF5 0.10bpac、YASS 设置 12 和 MME 0.10 上执行迭代最佳特性选择策略。结果如图 4 所示。个别图表显示了 OOB 的进展 k≤10 的误差估计以及所选子模型的列表。我们遵循图 1 中引入的子模型的符号，并通过灰色阴影和添加后缀“ref”来区分子模型的引用版本。作为比较，我们还包括使用整个 CC-JRM 时的 OOB 性能（最后一个独立离散的点）。

Figure 4 clearly demonstrates that it is indeed possible to obtain performance similar to CC-JRM with as few as one tenth of its submodels, reducing thus the testing time by one order of magnitude. The selected submodels are also generally different and algorithm-specific, which confirms our claim that there is no universally effective subset of CC-JRM.

The results provide us with another very interesting insight. Quite surprisingly, the appearance of a *reference* submodel often does not imply that the original version of the same submodel has been previously selected. In other words, a reference submodel may be useful as a complementary feature set to *other* types of features as well. Note, for example, the fourth selected submodel for YASS or the third for nsF5 and MME. This phenomenon indicates the intrinsic complexity of relationships among all extracted features and their reference values, and supports the hypothesis that appeared in [8]: individual features of complex feature spaces serve *each other* as references. For high-dimensional spaces, the concept of Cartesian calibration can thus be viewed simply as model enrichment that makes the feature space more diverse.

图 4 清楚地表明，只需要十分之一的子模型就可以获得与 CC-JRM 类似的性能，从而将测试时间减少一个数量级。所选的子模型通常也是不同的，并且是特定于算法的，这证实了我们的观点，即 CC-JRM 不存在普遍有效的子集。

批注 [.皓6]: 本来是 102 个子模型，现在运用 10 个子模型就可以得到 OOB 性能

这些结果为我们提供了另一个非常有趣的见解。相当令人惊讶的是，引用子模型的出现通常并不意味着相同子模型的原始版本已经被预先选择。换句话说，参考子模型也可以作为其他类型特性的补充特性集。注意,例如,选择第四子模型为 nsF5、YASS 或 MME 这一现象表明所有提取特性之间的关系的内在复杂性和参考价值,并支持的假设出现在[8]:单个特性的复杂功能空间相互服务引用。因此，对于高维空间，笛卡尔坐标校准的概念可以简单地看作是使特征空间更加多样化的模型丰富。

5. SUMMARY

Arguably, the most important element of today’s feature-based steganalyzers is the feature-space design. In this paper, we follow the recent trend of constructing rich feature spaces consisting of many simple submodels, each of them capturing different types of dependencies among coefficients. We constructed a rich model of JPEG images, abbreviated as CC-JRM, and demonstrated its capability to detect a wide range of qualitatively different embedding schemes. When combined with a scalable machine learning, CC-JRM outperforms all previously published models of JPEG images.

Merging the JPEG domain rich model with the spatial domain rich model (SRM) recently proposed in [4] results in a 35,263-dimensional features space that further improves steganalysis across all six tested steganographic schemes and all tested payloads. This confirms the thesis that steganalysis benefits from multiple-domain approaches.

The experiments from Section 4 indicate that the proposed CC-JRM does not contain any universally effective subset that could replace CC-JRM while keeping its performance across different stegoschemes. However, if we are to construct a targeted steganalyzer for detection of a selected steganographic method, it is possible to significantly reduce the dimensionality by supervised feature selection.

The last experiment of Section 4 showed that reference features are often useful even without their original feature values, which sheds more light on the real benefit of Cartesian calibration in high dimensions.

Matlab implementation of all feature sets used in this paper is available at http://dde.binghamton.edu/download/feature_extractors.

可以说，当今基于功能的隐写分析器最重要的元素是功能空间设计。在这篇论文中，我们跟随最近的趋势，**构造由许多简单的子模型**组成的丰富的特征空间，**每个子模型捕获不同类型的系数之间的依赖关系**。我们构建了 **JPEG 域富模型**，简称为 CC-JRM（隐写分析），并演示了其检测各种不同嵌入方案的能力。当与可伸缩的机器学习相结合时，CC-JRM 的性能优于所有以前发布的 JPEG 图像模型。

将 **JPEG 域富模型与最近在[4]中提出的空间域富模型(SRM)合并**，可以**得到 35263（J+SRM）维的特征空间**，从而进一步改进所有 **6 个**（nsF5、model-based steganography、YASS、MME、BCH、BCHopt）经过测试的隐**写方案**和所有经过测试的有效载荷的隐写分析。这**证实**了隐写分析从多域方法中获益的观点。

第 4 节的实验表明，所提出的 CC-JRM 不包含任何可以替代 CC-JRM 的通用有效子集（它的子集不能代替它），同时在不同的隐写器之间保持其性能。然而，如果我们要构建一个有针对性的隐写分析仪来检测所选择的隐写方法，则可以**通过监督特征选择（OOB 错误分析）来显著降低维数**。

第四部分的最后一个实验表明，即使没有原始的特征值，参考特征通常也是有用的，这进一步说明了高维笛卡尔坐标校准的实际好处。

本文使用的所有特征集的 Matlab 实现可以在 http://dde.binghamton.edu/download/feature_extractors 中找到。

1.

6. ACKNOWLEDGMENTS

The work on this paper was supported by Air Force Office of Scientific Research under the research grant number FA9550-09-1-0147. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation there on. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of AFOSR or the U.S. Government.