Detectability-based JPEG steganography modeling the processing pipeline: the noise-content trade-off

Quentin Giboulot, Rémi Cogranne, Patrick Bas

Detectability-based JPEG steganography modeling the processing pipeline: the noise-content trade-off

Quentin Giboulot[+], Rémi Cogranne[+], Patrick Bas[†]
[+]University of Technology of Troyes, France
[†]CNRS, CRIStAL Lab., École Centrale de Lille, France

***Abstract*—The current art of steganography shows that schemes using a deflection criterion (such as MiPOD) for JPEG steganography are usually subpar with respect to distortion-based schemes. We link this lack of performance to a poor estimation of the variance of the model of the noise on the cover image. However, this statistically-based method provides a better assessment of the detectability of hidden data as well as theoretical guarantees under a given model. In this paper, we propose a method to obtain better estimates of the variances of DCT coefficients by taking into account the dependencies introduced by development pipeline on pixels. A second method, which is a side-informed extension of Gaussian Embedding in the JPEG domain using quantization error as side-information, is also formulated and shown to achieve state-of-the-art performances. Eventually, the trade-off between noise and content complexity in steganography is thoroughly analyzed through the lenses of these two new methods using a wide range of numerical experiments.**

I. Introduction

Imperfect steganography rests on two pillars for its performance: the coding method and the cost function. The development of the Syndrome-Trellis Coding method [1], which is able to closely achieve the rate-distortion bound, led most of recent research efforts into designing better cost functions. There exists currently two approaches for designing such functions. The first, and most popular, is a heuristic approach, named distortion-based steganography. Here a cost function is heuristically defined to associate a cost to each pixel/DCT coefficient, the scheme then works by minimizing the sum of costs under the constraint of embedding a given payload. The current state-of-the-art in JPEG steganography, J-UNIWARD is a representative examples of this approach, as well as its more recent variants using filtering of costs [7]. J-UNIWARD costs are based on directional noise residuals estimated with a wavelet filter bank. Such costs are heuristically linked to local content complexity and the rationale is thus to assign high costs to smooth areas and low costs to areas which are "unpredictable" in every directions. Distortion-based schemes are thus often based on the concept of "content unpredictability". The major limitation of this approach is the absence of guarantees on the security of the steganography as the distortion function is not formally linked to any statistical detectability. In addition, the parameters of distortion function is often tailored to a specific detector and/or a specific dataset; as a consequence, the results may dramatically change when changing the dataset or the processing pipeline [2]. The second approach is the statistically-based one. It is based on minimizing the power of the most-powerful test under a given model of distribution for the cover and the stego. The costs associated to pixels/DCT coefficients are directly linked to how much the power of the detector would be increased if the pixel/DCT coefficient was modified. MiPOD [3] successfully used this approach in the spatial domain by modeling cover noise as realizations of independent Gaussian random variables and performing ternary embedding. More recently, the Gaussian embedding scheme [4] achieved state-of-the-art performance in the spatial domain using the same model but considering the stego-signal as an additive Gaussian noise instead. Contrary to distortion-based schemes, these algorithms will favor embedding in the noise of an image instead of the "content complexity" (though, see Section II.A, of [3] which explicitly takes the error of content estimation into account without further analysis on what form this error should take).

Despite such success in the spatial domain, the straightforward generalization of MiPOD to the JPEG domain was initially reported to have subpar security with respect to J-UNIWARD in [5][1]. As such, the only competitive steganographic alternative which relies on statistically defined costs in the JPEG domain uses so-called side-information. Side-information (SI) can refer to any knowledge, unknown to the steganalyst, on a given cover which can be used by the steganographer to improve the security of a steganographic scheme (e.g: another image of the same scene, the RAW image, knowledge of rounding errors, ...). In the case of JPEG images, SI usually relies on the so-called pre-cover which consists of the non-rounded DCT coefficients of the image. The state of the art for distortion-based scheme, SI-UNIWARD [6], significantly improves the security over J-UNIWARD by heuristically modulating its costs by $|0.5 - e_i|$ ($e_i = x_i - \mathrm{round}(x_i) \in]-0.5, 0.5]$ being the rounding error of the i-th DCT coefficient x_i), embedding preferentially in coefficients which are close to a bin boundary. On the other hand, model-based SI-MiPOD [5], referred to as MB-MiPOD in this paper, formally derives a modulation factor of $|0.5 - e_i|^2$ by minimizing the KL-divergence between the cover and the stego image conditioned on the realization of the pre-cover.

Recently, another trend developed in order to boost the practical security of embedding schemes, relies on the prin-

[1]We will, however, report contradicting results to those contained in [5]; the performance of JMiPOD being actually quite competitive, see also [28]

ciple of adversarial embedding using deep neural nets combined with appropriate retraining. The key idea of adversarial embedding [29] is (i) to build a classifier representing the adversary/steganalyst targeting a given steganographic scheme and payload, (ii) to modify the embedding scheme by tuning the embeddings costs in order to bypass the classifier. One can after retrain a new classifier and iterate. An efficient strategy to iterate [30] is to select the stego images which are the less detectable using the best trained classifiers. Note that, if the gain is substantial in term of undetectability after more than 5 iterations, this class of embedding schemes is heavy to deploy in practice. Indeed, the steganographer needs, in order to generate one stego image, first to generate several stego databases (one per iteration) and then to retrain a deep convolutional network at each iteration.

In [8], we linked the lack of performance of J-MiPOD to the fact that the estimation of the variances in the JPEG domain relied on the incorrect assumption of independence of the pixels in the spatial domain. Such a method leads to variances which are almost constant by blocks (up to the quantization step) in the JPEG domain. Taking into account the dependencies between pixels during the estimation of the variances leads to finer estimates. However, this paper will nuance this observation by showing that the empirical performance of such statistical based algorithm is also highly dependent on the presence of image content which is difficult to estimate. By only relying on the noise component in the image, such steganographic scheme can become sub-optimal when compared to distortion-based embedding schemes w.r.t. to empirical detectors, with limited possibilities for modeling covers, for a given dataset and steganalysis strategy. To alleviate this problem, we propose in this paper a methodology to add a specific component to the variance estimation which is only linked to the content of the image. The main contribution of this work is threefold:

1) We propose a multivariate statistical model that describes the distribution of DCT coefficients by estimating correlation between neighboring pixels after the development process. The model is used to provide better estimates of the noise variances.
2) A side-informed steganographic scheme in the JPEG domain is also proposed ; minimizing the power of the most powerful detector in the continuous domain, while the constraint of embedding a given payload length is expressed in the quantized domain.
3) Using these two methods, together with numerous different image datasets allows the understanding of the influence of the various parameters such as the trade-off between computing distortion based on content complexity and computing distortion based on noise variance.

The present paper is organized as follows:

1) The first part deals with the estimation of the covariances of DCT coefficients and mostly builds upon our previous work [8]. We propose a slight extension of the model to handle gamma corrections which is a common operation in image-processing pipeline which breaks both the stationarity and linearity assumptions of our model.
2) The second part deals with the side-informed Gaussian embedding scheme also proposed in our previous work. Following [3] we cast the problem as minimizing the power of the optimal detector under a payload constraint and we provide an analytic expression of the asymptotic power of the most powerful (MP) detector. Furthermore, we study the impact of the knowledge available to the steganalyst on the performance of the MP detector. Finally, we highlight the consequences of the independence assumption of the DCT coefficient on the performance of the likelihood ratio test as well as the performance of the empirical detectors.
3) The third part discusses the important choice of hiding locations for the steganographer, namely the choice of hiding behind content complexity or hiding behind the sensor noise. We show that this strategy highly depends on the training and testing dataset properties. We then propose a simple methodology to take into account content complexity independently of the noise in our proposed embedding method.
4) The fourth and last part presents and comments on the results of experiments and discuss the robustness of our method to different processing pipelines, datasets as well as the impact of noise and content complexity.

We would like to point out that the present paper is an extension our prior paper [8], in particular with respect to the statistical model of the cover and variance estimation method, Section II. The novelty of the present paper is the following:

- it extends the noise model to take gamma correction into account, providing a more comprehensive picture of processing pipeline,
- it analyzes the Gaussian embedding algorithm fully by deriving the MP detector as well as studying the relationship between the embedding probabilities and the quantization error,
- it proposes a method to take into account "content complexity" in order to deal with an imperfect Warden,
- it provides a much wider set of numerical results allowing, in particular, to discuss the trade-off between hiding information behind noise against hiding information behind content which is difficult to estimate for the steganalyst.

Note that for the rest of this paper, we make the unusual assumption that the steganographer, Alice, has access to the RAW image and to the processing pipeline from which her cover was generated. While access to a camera giving access to RAW files might have been difficult in the past, most cellphones today do give such an access, making this assumption believable even in a realistic setting.

The proposed scheme is also adversarial in the sense that the embedding is directly designed against an classifier, here the Likelihood Ratio Test. However contrary to adversarial embedding methods based on deep neural nets [29] combined with $\min\max$ iterations [30], our scheme is only based on the knowledge of the statistical properties of the develop-

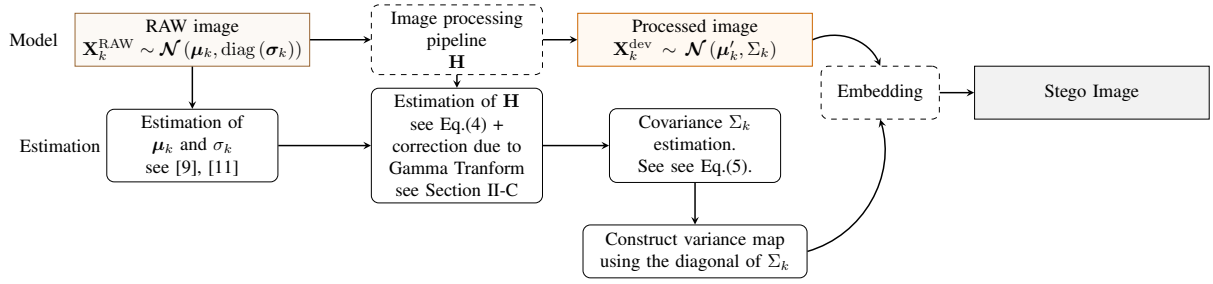


Fig. 1: Principle of steganography with side information coming from the processing pipeline: the estimation of the covariance matrix of the sensor-noise in the DCT domain enables to compute reliable variance estimates and to derive meaningful deflection coefficients, used to derive statistically founded costs for JPEG images.

ment pipeline, which can simply be estimated using only one (Raw,Cover) pair. Furthermore, it requires neither the computationally expensive training and retraining operations of [29], [30] nor the multiple generations of stego databases needed to update the adversary.

The following notational conventions will be used throughout the paper. Matrices and vectors will be typeset in boldface. When indexing, we use k to refer to vector elements, while i refers to individual scalar elements. Regarding specific notations, σ_i^2 and Σ_k always refer to the cover noise variance and covariance respectively. On the other hand ϵ_i refers to the stego-signal variance. $\mathbf{X}$ and its super-scripted and sub-scripted variants always refer to a cover element interpreted as a random variable, whereas $\mathbf{Y}$ always refers to a stego element interpreted as a random variable. Finally z_i always refers to the realization of an image element when it is not known beforehand if this image element comes from a cover or stego image.

II. Statistical Model of the Sensor Noise in the JPEG Domain

In this paper we only consider cover images in DCT (before rounding) and JPEG (after rounding) domains. The methodology proposed in the present paper is based on a statistical model of DCT coefficients corrupted by multivariate and correlated Gaussian noise in the spatial domain. In addition, we assume that the image used to hide data was captured by the steganographer. Alice thus has access to both the raw image and to the processing pipeline which generates the cover from the RAW. In this section, we are only interested in the specification and estimation of the noise model of the cover image. To facilitate both the specification of the model as well as its estimation, it is assumed that the processing pipeline (1) can be approximated by a linear operator on blocks of pixels/DCT coefficients (linearity assumption), and (2) is identical over all the blocks of the image (stationary assumption).

A. Model Definition

Our goal is now to produce a model which allows to take correlations between DCT coefficients into account in order to obtain high quality estimates of the variance for each DCT coefficient.

To do so requires first a model of the processing pipeline. To develop one 8×8 block of photo-sites, a processing pipeline will usually need to interpolate photo-sites inside this block with photo-sites outside of it. An important example is bilinear demosaicking which, for a photo-site of given color, uses its neighbors to produce the interpolated value of a different color – see Figure 2. This creates correlation between neighboring photo-sites. Furthermore the DCT transform at the end of the pipeline will propagate such correlations to the entirety of direct neighboring blocks. For this reason, our model of the processing pipeline will need as input blocks of size at least $(8 \times M)^2$ – with $M \in \mathbb{N}$ – where M^2 is the number of input block necessary for the processing pipeline to be able to develop **one** 8×8 block of DCT coefficients. Note that, for the reasons just invoked, in the case of a linear pipeline comprised of bilinear demosaicking, RGB to grey conversion and DCT transform, we would need nine 8×8 blocks of photo-sites to obtain one 8×8 block of DCT coefficient (the block of photo-site itself as well as all its neighboring blocks).

For some pipeline, adding neighboring blocks might not be sufficient to be able to compute all correlations of interest. For example, in the case of bilinear demosaicking, we must add a margin of photo-sites at each border of the "macro-block" of $(8 \times 3)^2$ photo-sites to be able to compute the correct value of the correlations at the borders of macro-block.

Concluding this discussion, we model the processing pipeline up to and including the DCT transform (i.e. demosaicking, white balancing, denoising, etc…) as a linear operator represented as a matrix $\mathbf{H}$ of dimension $8^2 \times (8M+m)^2$ where m is the size of the margin at each border of the $(8M+m)^2$ "macro-block". We next go on to define the model of the sensor noise.

Following [9], [11], we model photo-site values as independent random variables drawn from the following heteroscedastic noise model:

$$X_i^{\mathrm{RAW}} \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right) \;;\; \sigma_i^2 = c_1 \mu_i + c_2, \tag{1}$$

where μ_i is the value of the i-th photo-site that would have been observed if it had not been corrupted by acquisition noise. c_1 and c_2 are the parameters of the heteroscedastic model which depend on the sensor and the ISO to capture the image. For simplicity and clarity, we will model a block of $(8M+$

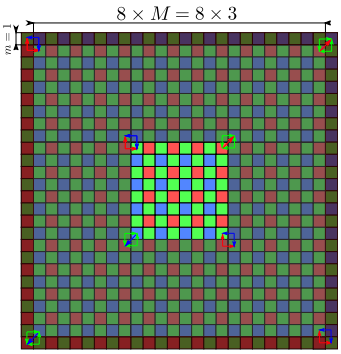


Fig. 2: Dependency structure for a linear pipeline. Due to the use of bilinear demosaicking, adjacent 8-connected block of 8×8 photo-sites will be correlated after the operation. To be able to take into account all correlations in the DCT domain, a margin of one row/column is necessary at every border of the blocks to compute the correct value of the correlations of the neighboring blocks after demosaicking. Notice that the margin is sufficient because the central block is independent from the block containing the margin.

$m) \times (8M + m)$ photo-sites jointly by:

$$\mathbf{X}_k^{\mathrm{RAW}} \sim \mathcal{N}(\boldsymbol{\mu}_k, \mathrm{diag}(\boldsymbol{\sigma}_k)), \tag{2}$$

where $\mathbf{X}_k^{\mathrm{RAW}}$ corresponds to the $(8M + m) \times (8M + m)$ block of photo-sites centered on the k-th 8×8 block of the RAW image. By $\mathrm{diag}(.)$ we refer to the diagonal covariance matrix generated by a vector of variances.

Since the processing pipeline introduces dependencies between pixels, and since we suppose it to be linear, we can model 8×8 blocks of dependent DCT coefficients of the developed image as multivariate Gaussian random variables [12]:

$$\mathbf{X}_k^{\mathrm{dev}} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{\mu}'_k, \Sigma_k\right) \ ; \ \Sigma_k = \mathbf{H}\left(\mathrm{diag}(\boldsymbol{\sigma}_k)\right)\mathbf{H}^T. \tag{3}$$

It is important to note that no assumption of independence is made at this point between neighboring 8×8 block of DCT coefficient. However, if one is interested in modeling inter-block dependencies directly, the model can easily be generalized by setting the output size of $\mathbf{H}$ as $(8N)^2$ with $N^2 < M^2$ being the number of dependent neighboring blocks to be modeled for each 8×8 block of DCT coefficient.

B. RAW model Estimation and Covariance Estimation

To be able to estimate the covariances Σ_k, we need to estimate the variances associated to each photo-site which are themselves function of the expected value of each photo-site μ_i, see Eq. (1). To that end, we denoise the RAW image using the method in [13] based on the inverse Anscombe transform and the BM3D algorithm [14]. The heteroscedastic model parameters c_1 and c_2 which characterize the noise variance as a function of the expectation Eq. (1), are then estimated using the method detailed in [9], [11]. In practice, the RAW model only needs to be estimated once for each camera and each ISO among images in a given dataset. Once the parameters of the RAW model – c_1 and c_2 – are estimated, the covariance matrix of each block Σ_k is simply given by:

$$\Sigma_k = \mathbf{H}\left(\mathrm{diag}(\boldsymbol{\sigma}_k)\right)\mathbf{H}^T. \tag{4}$$

since $\mathbf{H}$ is considered to be a linear operator.

Depending on the processing pipeline, $\mathbf{H}$ can be computed analytically such as in [15] for bilinear demosaicking. In our case, we adopt a different approach by approximating the processing pipeline as a stationary linear operator. This approximation may be slightly less accurate, however, it comes with great simplicity and a much broader scope of application since several processing may be very difficult to model or even be partially unknown (such as denoising or complex demosaicking). We propose to estimate $\mathbf{H}$ blindly using a least square estimation between the photo-site blocks and the developed blocks, that is, solving for $\mathbf{H}$:

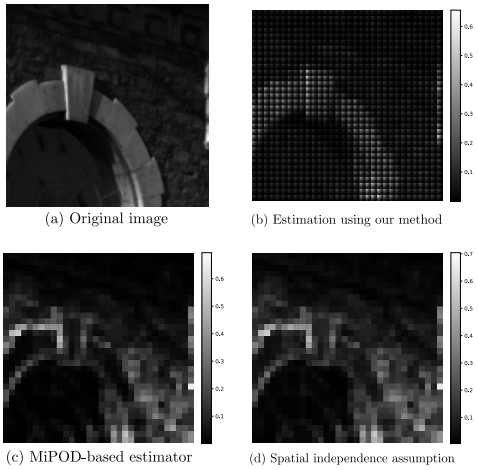
$$\mathbf{X}_k^{\mathrm{dev}} = \mathbf{H}\mathbf{X}_k^{\mathrm{RAW}}, \tag{5}$$

with the least square solution being:

$$\mathbf{H} = \mathbf{X}_k^{\mathrm{dev}}\left(\mathbf{X}_k^{\mathrm{RAW}}\right)^T\left(\mathbf{X}_k^{\mathrm{RAW}}\left(\mathbf{X}_k^{\mathrm{RAW}}\right)^T\right)^{-1}. \tag{6}$$

Even though this method allows the estimation of the full covariance matrix, the embedding scheme designed in Section III will only make use of its diagonal. Indeed, the design of a non-addtive scheme being out of the scope of this paper, we consider the DCT coefficient to be independent during the embedding phase. The reader might then rightly ask the point of going to the trouble of considering the DCT coefficients to be dependent until the end of the pipeline to then forego this assumption altogether. The reason is that it allows a far better estimation of the variance in the JPEG domain than, say, computing directly the variances in the spatial domain, effectively considering pixels to be independent and then propagating those variances in the JPEG domain using the DCT transform.

To illustrate this point, we show the variance maps estimated with and without keeping the dependencies between pixels until the end of the estimation in Figure 3. First, observe that when using our model (upper and lower right figure), the variance of the noise is directly related to the amplitude of the signal – this is a consequence of the heteroscedastic model underlying our method. On the other hand, when using a method based on MiPOD's estimator which is sensitive to content – see Section IV – the variance is no more dependent on the amplitude of the signal but to the presence of sharp transitions and texture. Second, we highlight the importance of using correlations until the end of the variance estimation. Indeed, observe that when using our method as described, the variance decreases as the DCT mode increases hence correctly taking account the low-pass effect of the JPEG compression (and here of the bilinear demosaicking). On the other hand, if we assume pixels to be independent, like in the lower-left and lower-right figures, we observe that the variance is almost constant by block of DCT coefficient.



(a) Original image (b) Estimation using our method

(c) MiPOD-based estimator (d) Spatial independence assumption

Fig. 3: Comparison of the different structure of the variance maps of the DCT coefficients of image *5105* of the BOSS dataset processed using the linear pipeline (see Table IV) at QF100 depending on the estimation method used. The variance map is either estimated using the method outlined in Section II (upper-right corner) or by assuming pixels to be independent (lower right and lower left). The lower-left figure is obtained by using the variance estimator define in Section IV based on MiPOD's estimator and propagating the variances in the DCT domain using Eq (29). The lower-right one is obtained using our method to estimate variances in the **spatial domain** but then foregoing dependencies between pixels by propagating variances in the DCT domain in the same way as for the MiPOD-based estimator.

C. Gamma Correction

Along the general image processing pipeline, there is one step that can hardly be approximated as a stationary linear operator, namely the gamma correction. Indeed, this compensation for the color response function of monitors is both non-linear and non stationary. More specifically, gamma correction is a piecewise-continuous function with one linear part and a non linear part. Its standard form (as used in rawpy/libraw for example) is given by

$$\Gamma(x) = \begin{cases} \gamma_0 x, & \text{if } x < 0.018, \\ 1.099 x^{\frac{1}{\gamma_1}} - 0.099, & \text{if } x \geq 0.018, \end{cases} \tag{7}$$

where γ_0 and γ_1 are specified by the user and $x \in [0,1]$ is the normalized input pixel value.

Since this operation is usually applied after demosaicking, on each colour channel separately, the estimated variance must be corrected at this step of the processing to be able to estimate correctly the variance further up in the processing pipeline. To that end, we propose to use the Delta method [16, Theorem 11.2.14] to compute the covariance matrix using a first order Taylor approximation around the mean:

$$\boldsymbol{\Sigma}_k^{\Gamma} \doteq \nabla\Gamma\left(\boldsymbol{\mu}_k^{\mathrm{RGB}}\right) \boldsymbol{\Sigma}_k^{\mathrm{RGB}} \nabla\Gamma\left(\boldsymbol{\mu}_k^{\mathrm{RGB}}\right)^T. \tag{8}$$

where $\nabla\Gamma$ is the gradient of the gamma correction function, $\boldsymbol{\mu}_k^{\mathrm{RGB}}$ and $\boldsymbol{\Sigma}_k^{\mathrm{RGB}}$ are respectively the mean value and covariance matrix of the k-th block of of pixel before gamma correction.

Note that through this approximation, we suppose the transform to be linear around the mean ; hence effectively approximating the pixel noise value to follow a normal distribution even after gamma correction. Also note that this approximation necessitates that the mean, hence the denoised version of the demosaicked image before gamma correction, to be computed by the steganographer. In this paper we used the BM3D denoising algorithm [10] to denoise the RAW image. Once the denoised demosaicked image is computed, the covariance matrix of the blocks of the demosaicked image must be computed for each color channel before being corrected using Eq. (8).

III. Embedding

We have described in the previous section how to estimate the covariance matrix of DCT coefficients. Once this task is carried out, the main problem for the steganographer remains to associate to each pixel a relevant probability of embedding in order to minimize detectability. Such a method has been proposed in [3]. In the present paper, a different, original method is used to avoid several difficulties. In particular, the proposed scheme, which we will call **SI-Gaussian** for the rest of this paper, does not rely on the fine quantization assumption; such an assumption is not relevant for DCT coefficient of JPEG images. The proposed scheme also takes into account the side-information from the non-quantized value of DCT coefficients in a natural way without any ad-hoc modulation of the costs.

A. Cover and Stego Image Model

While the model presented in Section II would allow to consider the dependencies between DCT coefficients, this would require the design of a distortion function able to consider DCT coefficients jointly. However, the design of a non-additive scheme is out of the scope of this paper and is left out for future research. We will thus model the pre-cover as a N-dimensional vector of DCT coefficients $\mathbf{X} = (x_1, x_2, \ldots, x_N)$ considered as realization of N independent Gaussian variables:

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i) \tag{9}$$

The idea is to minimize the impact of the embedding directly in the continuous domain in order to:

1) Take the side-information into account,
2) Relax the fine quantization limit assumption.

To that end, we add a pre-stego signal to the pre-cover as a zero-mean Gaussian signal with variance ϵ_i^2 leading to a pre-stego content modeled as a N-dimensional vector $\mathbf{Y} = (y_1, y_2, \ldots, y_N)$ which is made of realizations of N independent Gaussian variables:

$$Y_i \sim \mathcal{N}\left(\mu_i, (\sigma_i^s)^2\right) \ ; \ (\sigma_i^s)^2 = \sigma_i^2 + \epsilon_i^2. \tag{10}$$

When using the variance estimation method given in Section II, one has to pass from the multivariate model to the independent model given here. Since only the variances σ_i are of interest, this is simply done by keeping only the diagonal of the estimated covariance matrices: the l-th DCT coefficient contained in the k-th 8×8 block is thus assigned the variance at the (l,l)-th position of the Σ_k matrix.

B. Optimal test

Following MiPOD's methodology, we design a steganographic scheme which generates a signal that minimizes the power of the most powerful (MP) detector. To cast the steganography problem into the continuous domain, we will work under the assumption that the Warden knows the stego signal variance $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_N)$ as well as the model parameters $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ and analyses the image before rounding $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)$. The Warden's goal is to decide between the two hypotheses $\forall i \in \{1, 2, \ldots, N\}$:

$$\begin{cases} \mathcal{H}_0 & = \left\{ z_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right) \right\}, \\ \mathcal{H}_1 & = \left\{ z_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2 + \epsilon_i^2\right) \right\}. \end{cases} \tag{11}$$

Set the pdf of the noise distribution under $\mathcal{H}_0$, $p_{\sigma_i}(x)$ and $q_{\sigma_i,\epsilon_i}(x)$ under $\mathcal{H}_1$ as

$$p_{\sigma_i}(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(x-\mu_i)^2}{2\sigma_i^2} \right) \tag{12}$$

$$q_{\sigma_i,\epsilon_i}(x) = \frac{1}{\sqrt{2\pi(\sigma_i^2+\epsilon_i^2)}} \exp\left(\frac{-(x-\mu_i)^2}{2(\sigma_i^2+\epsilon_i^2)} \right). \tag{13}$$

We can then use the Neyman-Pearson criterion of optimality. In this case the Warden constructs a test $\delta : \mathbb{R} \to \{\mathcal{H}_0, \mathcal{H}_1\}$ which maximizes the power of the test $P_D \triangleq \mathbb{P}(\delta(x) = \mathcal{H}_1 | \mathcal{H}_1)$ under a given false-alarm probability $P_{FA} \triangleq \mathbb{P}(\delta(x) = \mathcal{H}_1 | \mathcal{H}_0)$.

Under these assumptions, the problem of the Warden (11) is reduced to a choice between two simple hypotheses for which the Neyman-Pearson Lemma states that the most-powerful test is the likelihood ratio test (LRT), defined, in our case as follows:

$$\Lambda_i(z, \sigma_i, \epsilon_i) = \ln\left(\frac{p_{\sigma_i}(z)}{q_{\sigma_i,\epsilon_i}(z)} \right), \tag{14}$$

$$\Lambda(\mathbf{z}, \sigma, \epsilon) = \sum_{i=0}^{N} \Lambda_i(z_i, \sigma_i, \epsilon_i) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \tau, \tag{15}$$

since we assume independence between DCT coefficients.

We show in the Appendix that the asymptotic power of the LRT as the number of DCT coefficient $N \to \infty$ is given by :

$$P_D = \mathbb{P}(\delta(x) = \mathcal{H}_1 | \mathcal{H}_1) \tag{16}$$

$$= Q\left(\frac{Q^{-1}(P_{FA}) \sqrt{\mathrm{Var}_{\mathcal{H}_0}[\Lambda]} + \mathbb{E}_{\mathcal{H}_0}[\Lambda] - \mathbb{E}_{\mathcal{H}_1}[\Lambda]}{\sqrt{\mathrm{Var}_{\mathcal{H}_1}[\Lambda]}} \right), \tag{17}$$

$$= Q\left(\frac{Q^{-1}(P_{FA}) \sqrt{\sum_{i=1}^{N} \frac{\epsilon_i^4}{2\left(\epsilon_i^2+\sigma_i^2\right)^2}} + \sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4+2\sigma_i^2\epsilon_i^2}}{\sqrt{\sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4}}} \right) \tag{18}$$

where $Q(\cdot)$ is the tail distribution function of the standard normal distribution, with each of the moments derived in the Appendix. Under the assumption that the power of the stego signal is negligible compared to the sensor noise, that is $\sigma_i^2 >> \epsilon_i^2$, we obtain the much more mangeable expression:

$$P_D \doteq Q\left(Q^{-1}(P_{FA}) + \sqrt{\sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4}} \right). \tag{19}$$

C. Embedding by Minimizing the Power of the MP Detector

From Eq. (41) in the Appendix, one can note that the power of the MP detector is specified by two quantities: $\frac{\mathbb{E}_{\mathcal{H}_0}[\Lambda] - \mathbb{E}_{\mathcal{H}_1}[\Lambda]}{\sqrt{\mathrm{Var}_{\mathcal{H}_1}[\Lambda]}}$ and $\frac{\mathrm{Var}_{\mathcal{H}_0}[\Lambda]}{\mathrm{Var}_{\mathcal{H}_1}[\Lambda]}$. The first quantifies the contribution of the shift between the two Gaussian while the second quantifies the contribution due to the spread of the Gaussian.

To simplify the optimization problem, it is usual to neglect the second quantity and consider the variances of the LR equal under both hypothesis (see the Appendix in [3] and Section 2 in [4]). Under this approximation, which essentially consists in assuming that the stego-signal has a negligible power when compared to the natural image noise, we have $\sigma_i^2 >> \epsilon_i^2$ and get:

$$\begin{aligned} \frac{\mathbb{E}_{\mathcal{H}_0}[\Lambda] - \mathbb{E}_{\mathcal{H}_1}[\Lambda]}{\sqrt{\mathrm{Var}_{\mathcal{H}_1}[\Lambda]}} &= \frac{\sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4+2\sigma_i^2\epsilon_i^2}}{\sqrt{\sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4}}}, \\ &\doteq \sqrt{\sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4}}. \end{aligned} \tag{20}$$

Interestingly, the term $\varrho \triangleq \sqrt{\sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4}}$ entirely characterizes the power of the most-powerful test. This factor is thus the deflection coefficient for detectability of the best Warden one can face.

One can also note that, under the present method, the deflection only depends indirectly on the payload since we add a stego-signal whose variance should ensure that the embedding payload can be reached. This is highlighted in the following optimization problem the steganographer seeks to solve:

$$\begin{cases} \min\limits_{\epsilon_i} \varrho^2 & = \sum\limits_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4}, \\ M & = \sum\limits_{i=0}^{n} \sum\limits_{k \in \mathbb{Z}} \beta_i^k \log\left(\beta_i^k\right), \end{cases} \tag{21}$$

with

$$\beta_i^k = \mathbf{\Phi}\left(\frac{k - r_i + 0.5}{\epsilon_i} \right) - \mathbf{\Phi}\left(\frac{k - r_i - 0.5}{\epsilon_i} \right), \tag{22}$$

being the probability of modifying the i-th coefficient by $+k$, ,$\mathbf{\Phi}(\cdot)$ being the cumulative distribution function of the standard normal distribution, and $r_i = x_i - [x_i]$ being the rounding error of i-th DCT coefficient. In practice, the alphabet size of the embedding scheme must be finite, k is thus constrained to a finite range and the β_i^k renormalized accordingly.

To compute the β_i^k, we follow the same strategy as in [4]. Using the differential entropy of the Gaussian distribution and

the method of Lagrange multipliers, we can express the ϵ_i as a function of λ:

$$\mathcal{L} = \sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4} + \lambda \left(M^* - \frac{1}{2} \sum_{i=0}^{N} \log\left(2\pi e \epsilon_i^2\right) \right). \tag{23}$$

Solving:

$$\frac{\partial \mathcal{L}}{\partial \epsilon_i} = 0 \iff \frac{2\epsilon_i^3}{\sigma_i^4} + \frac{\lambda}{\epsilon_i} = 0, \tag{24}$$

we obtain:

$$\epsilon_i^2 = \sqrt{\frac{\lambda}{2}} \sigma_i^2. \tag{25}$$

Finally, we obtain the solution for λ:

$$\lambda = \exp\left(\frac{2M^*}{N} - \frac{4}{\mathrm{N}} \sum_{\mathrm{i=0}}^{\mathrm{N}} \log\left(\sigma_{\mathrm{i}}\right) - 2\log(2\pi\mathrm{e}) + \log(2) \right). \tag{26}$$

Note that we use a proxy parameter M^* which is the **entropy in the continuous domain**. The goal is to find the entropy in the continuous domain which leads to the desired entropy in the discrete domain. Hence, the problem reduces to finding M^* such that $M = \sum_{i=0}^{n} \sum_{k \in \mathbb{Z}} \beta_i^k \log\left(\beta_i^k\right)$ (note that M is a function of M^*). This problem can be solved through a binary search over M^*.

Looking at the system in Eq. (21), one should observe that the rounding error is not directly taken into account in the quantity to minimize; it is only taken indirectly into account in the constraint as values closer to the bin boundary will contribute more to the total entropy.

To understand how the trade-off between the side information and the variance emerges from the optimization, we plot in Figure 4 the embedding probabilities β_i as a function of both the rounding error r_i and the variance σ_i^2, this is obtained for a 256×256 image with variances taking value in $\{0.1, 10\}$ with an embedding with payload $M = 0.25$ bpp. We compare the embedding probabilities obtained with our method denoted "SI-Gaussian", with those obtained with a side informed version of MiPOD where MiPOD's costs are modulated with the so called "Minimum Perturbation" heuristic – see [18] – which is currently the heuristic giving the best security performance when using side-information.

Be it for SI-Gaussian or SI-MiPOD, the probability of embedding in a coefficient increases w.r.t both r_i and σ_i^2. However, one can observe that our method is more sensitive to both these 2 parameters than SI-MiPOD. Indeed, coefficients with low variances will only be used when they are almost at the bin boundary while coefficients with higher variances will always be used more often when compared to SI-MiPOD. This behavior emerges from the optimization system as follows: since the objective function is expressed in the continuous domain, the variance of the stego-signal ϵ_i grows linearly with the variance of the coefficient σ_i (Eq. (25)). As such the only effect of the constraint is to increase or decrease λ until the constraint is met. For reasonable payloads, this translates to having very low stego-signal variances for the lowest coefficient variances and, correspondingly, high stego-signal variance for coefficients with the highest variance. As a consequence, coefficients with low variances only cross the bin boundary if they are already on the boundary while those with high variances cross the boundary even at value closer to the center.

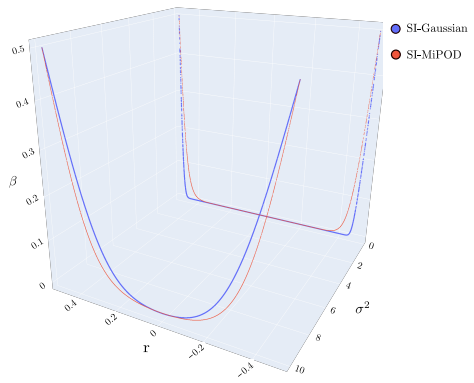


Fig. 4: β_i as a function of the rounding error r_i and of the variance σ_i^2 on an 256×256 image with variances taking value in $\{0.1, 10\}$ with an embedding with payload $M = 0.25$ bpp. The red dots correspond to embedding probabilities obtained with SI-Gaussian while the red dots correspond to those obtained using SI-MiPOD with cost modulated using the Minimum Perturbation heuristic.

D. Influence of the Warden Knowledge

In this section we study the influence of the Warden knowledge about the cover on the performance of the LRT. This study allows validating our choice of minimizing the power of the MP detector directly in the continuous domain. We also provide an interpretation of the success of side-information in terms of impossibility of estimating the selection-channel.

The core design of the Gaussian embedding method proposed in Section III-C, Eq. (21), essentially assumes that the steganalyst is able to access both the unquantized version of the image and a perfect estimate of the variances of the DCT coefficients σ_i^2 and the variance of the stego-signal ϵ_i^2. This means that we also implicitly assume that the Warden does not have access to the original rounding errors as this would render the detection easier since the embedding probabilities directly depend on them.

Following a similar presentation as in [3, Sec.III.A], we distinguish three possible kinds of Warden[2]:

1) An **Omniscient Warden** who knows the covariance matrix of DCT coefficients as well as the unquantized version of the sample and the variances of the stego-signal. As opposed to the embedding model used by the steganographer which for simplicity neglects the correlation between DCT coefficients, see Section II, the omniscient Warden has the advantage of knowing those

[2]Note that contrary to [3], we assume that all three Warden know the payload size.

correlations and can leverage this knowledge during the detection. To facilitate the computation of the LRT, this Warden considers only intra-block dependencies.

2) A **Knowledgeable Warden** who has access to the unquantized version of the sample, the variances of the DCT coefficient and the variances of the stego signal.
3) An **Ignorant Warden** who has only access to the quantized version of the sample along with the variances of the DCT coefficient. The ignorant Warden has thus no information on the stego-signal and thus can only estimate the selection channel through variances of the DCT coefficient since she does not know the rounding errors. In this situation, we will suppose that the Warden uses MiPOD's model to estimate the embedding probabilities β_i from the variances – see [3, Section 2].

We summarize the information to which each Warden has access to as well as the test they each perform in Table I. The LRT of the knowledgeable and ignorant Warden is computed empirically for each image using:

$$\Lambda(\mathbf{z}) = \sum_{i=1}^{N} \ln\left(\frac{p_{\mathcal{H}_1}(z_i)}{p_{\mathcal{H}_0}(z_i)}\right) \tag{27}$$

where z_i is the i-th- DCT coefficient while $p_{\mathcal{H}_0}$ and $p_{\mathcal{H}_1}$ are the pdf of the distribution under each hypotheses as given in Table II. For the omniscient Warden, the formula is identical except we use blocks of of DCT coefficients:

$$\Lambda(\mathbf{z}) = \sum_{k=1}^{N_{\text{block}}} \ln\left(\frac{p_{\mathcal{H}_1}(\mathbf{z}_k)}{p_{\mathcal{H}_0}(\mathbf{z}_k)}\right) \tag{28}$$

Note that we suppose that every Warden has access to the nuisance parameter μ_i.

To test the performance of each Warden, we took one RAW image from E1Base (*IMG_3699.cr2*) and simulated 10 000 noisified versions of this RAW image by adding an heteroscedastic noise to each photo site with $c_1 = 1.3$ and $c_2 = -7500$. The images were then demosaicked using the bilinear demosaicking algorithm, converted to greyscale, cropped to 256×256 at an offset of 2048 in each direction and finally compressed to JPEG with quality factor 100. Images were embedded with the method described in Section III-C at 0.15bpp. The LRT was then computed on each image for each different Warden. We also compared the LRT to three empirical detectors using either DCTR [19], GFR [26] or CCJRM [25] feature sets, coupled with the Linear Low Complexity Classifier (LCLC) [20] Figure 5 shows the result. SRNet [27] was also tested but its ROC curve was indistinguishable from the Omnisicent Warden at the payload of interest. The fact that SRNet is able to match the Omniscient Warden hints that it is somehow able take correlations of the DCT coefficients into account. This observation should be taken with caution since all images in the dataset have been generated from the same images. The estimation of a covariance matrix by the steganalyst is far less amenable on a more realistic dataset, where every image might have a different processing pipeline and acquisition parameters.

As expected, Figure 5, clearly shows that the knowledgeable Warden (orange full line) performs better than the ignorant

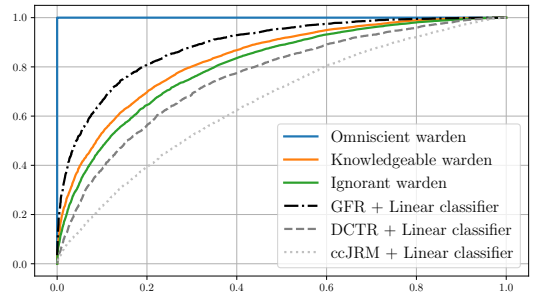


Fig. 5: ROC curves of the LRT for the different types of Warden/models compared with the empirical detector using DCTR and LCLC on 10000 noisified version of the RAW image *IMG_3699.cr2* from E1Base. The resulting noise is correlated due to the processing pipeline. Images are compressed with JPEG at QF100 embedded with Σ-SI-Gaussian at 0.15bpp.

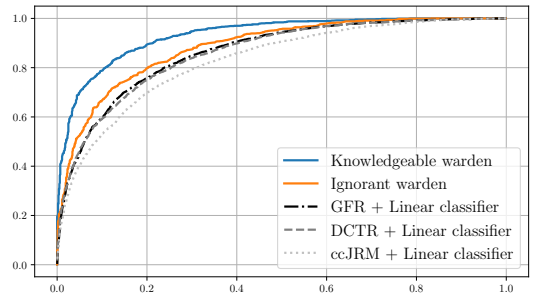


Fig. 6: ROC curves of the LRT for the different types of Warden/models compared with the empirical detector using DCTR and LCLC on 10000 noisified version of the JPEG from *IMG_3699.cr2* from E1Base. The resulting noise is independent. Images are compressed with JPEG at QF100 embedded with Σ-SI-Gaussian at 0.15bpp.

Warden (green full line) due to a better access to the selection channel. Interestingly, the ROC curves show that our scheme has no security if the covariances are available to the steganalyst, showing that correlations between DCT coefficients are important; taking such correlations into account could actually greatly improve the security of a steganographic scheme.

As for empirical detectors DCTR and cc-JRM perform worse than the ignorant Warden by 3% and 11% in terms of P_E respectively. On the other hand, surprisingly, GFR beats both the knowledgeable and ignorant LRT by 5% and 7% respectively. This shows that GFR is the only state-of-the-art rich model able to leverage dependencies between DCT coefficients to improve detection. The case of cc-JRM is interesting because even though it is built using co-occurrences of histograms, hence explicitly trying to take dependencies of DCT coefficients into account, it is still far worse than DCTR. Furthermore, GFR is built similarly to DCTR with first order statistics; this would tend to show that the choice of filter bank plays a more important role than the statistics themselves when trying to take correlations between DCT coefficients into account.

To validate our hypothesis that the observed gain in detec-

TABLE I: Summary of the information accessible to each Warden

Access to \Warden	**Omniscient**	**Knowledgeable**	**Ignorant**
Unquantized samples z_i	Yes	Yes	No – quantized only
Noise covariances Σ_k	Yes	No – variances σ_i^2 only	No – variances σ_i^2 only
Stego variances ϵ_i^2	Yes	Yes	No

TABLE II: Description of the model used by each warden

Models \Warden	**Omniscient**	**Knowledgeable**	**Ignorant**
$\mathcal{H}_0$	$\boldsymbol{z}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k)$	$z_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$	$[z_i] \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$
$\mathcal{H}_1$	$\boldsymbol{z}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k + \mathrm{diag}(\boldsymbol{\epsilon}_k))$	$z_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2 + \epsilon_i^2\right)$	$[z_i] \sim (1-2\beta_i)\mathcal{N}\left(\mu_i, \sigma_i^2\right)$ $+ \beta_i \mathcal{N}\left(\mu_i + 1, \sigma_i^2\right)$ $+ \beta_i \mathcal{N}\left(\mu_i - 1, \sigma_i^2\right)$

tion of GFR compared to DCTR is indeed due to the fact that GFR is able to take correlations between DCT coefficients into account, we repeated this experiment using independent noise. To do so, we first denoised the RAW image *IMG_3699.cr2* using BM3D, developed it using the same processing pipeline as previously and added independent Gaussian noise to each DCT coefficient with variances estimated on the noisy developed *IMG_3699.cr2* image using the method in Section II. The rest of the experiment was performed identically as the previous one. Results are presented in Figure 6.

These results validate our hypothesis as, this time, the three empirical detectors are in fact quite close in terms of performance (with cc-JRM lagging somewhat behind). Furthermore, all three of them are worse than both the ignorant and knowledgeable Warden [3]. We can conclude that since GFR is unable to use the correlations, its performance can't improve with respect to the other detectors and hence that the difference w.r.t. DCTR mostly lies in this information captured by the Gabor filters but not by the DCT modes.

IV. Trade-off Between Noise and Content Complexity in Steganography

Under the model presented in the previous section, only two parameters are of interest to the steganographer : the mean vector and covariance matrix of each individual block of DCT coefficients as they fully specify the model. However, since the steganalyst has only access to the stego image in the JPEG domain, her estimation of the model will necessarily be worse than the estimation of the steganographer.

This fact means that the steganographer can improve the security of his stego scheme by taking into account the estimation error of the model by the steganalyst. This observation was actually one of the cornerstone of MiPOD for which the authors showed that MiPOD's security could actually be improved by using subpar variance estimators.

In the previous section, we only tried to be optimal against the MP detector which might thus be sub-optimal when facing a practical detector. We shall see in Section V that building the MP detector considering only the photonic noise, and not taking into account image content, can decrease the practical security of the proposed embedding method. More specifically, using only the photonic noise is sub-optimal compared to the UNIWARD family which was specifically designed to embed in image locations where the estimation of the model parameters is difficult for the steganalyst.

We hypothesize that this loss of performance is due to the fact that, under generally adopted steganalysis settings, the learned classifier has difficulties to estimate the cover component. Whenever a large dataset of images like BossBase is used for training a classifier, the classifier can only rely on so called "inter-dependencies" between images to reduce the impact of imprecise content estimation for images which share a common source and common content. It is thus expected that a given dataset will favor different trade-offs between hiding information into the noise or into the content depending on the inter and intra-dependencies it contains.

In this section we consequently propose a simple methodology to take the content of the image into account.

A. Variance of the Estimation Error Related to the Content

We propose to estimate the variance that would be estimated on the image if no noise is present. Following what is proposed in [3, Section IIA] we model the variance of a DCT coefficient simply as the sum of the noise variance σ_i^2 as described in section II and the variance due to the modeling error ξ_i^2. This normality of the estimation is here assumed for computational tractability and ease of mathematical statement.

To estimate ξ_i^2, we work in the spatial model before the DCT transform and perform the following operations:

1) Estimate the pixel noise variances in the spatial domain using the method in Section II.
2) Denoise the image using the wavelet denoising algorithm[4] outlined in [21] using the estimated noise variances.
3) Apply MiPOD's variance estimation method on the denoised image, that is a Wiener filter with a window width of 2 followed by fitting of a parametric model with two-dimensional trigonometric polynomials.

[3]Note that the omniscient and knowledgeable Wardens are identical for the case of independent noise.

[4]We also experimented using BM3D for the denoising but the wavelet denoising algorithm consistently led to better security performances while also embedding faster by several orders of magnitude.

4) Compute the corresponding variance in the DCT domain using Eq. (29):

$$(\xi_{k,l}^{(a,b)})^2 = \sum_{i,j=0}^{7} (f_{i,j}^{k,l})^2 \cdot (\xi_{i,j}^{(a,b)})^2 / q_{l,k}^2, \tag{29}$$

with $f_{i,j}^{k,l} = \frac{1}{4} w_k w_l \cos\frac{\pi k(2i+1)}{16} \cos\frac{\pi l(2j+1)}{16}$, $w_0 = \frac{1}{\sqrt{2}}$ and $w_k = 1$ for $k > 0$.

Once the variance due to content is estimated, we finally sum the two variances for each DCT coefficient, σ_i^2 corresponding to the noise component and ξ_i^2 corresponding to the modeling error – that is the content complexity – component. The embedding using this variance map is then performed as usual. This way of computing the variances can be seen as a trade-off between hiding information both behind the photonic noise and behind the content complexity. Hiding only behind noise postulates a detector which can estimate the content of the image – the mean value of the DCT coefficient – perfectly. However, the current art in steganalysis is still sensitive to content complexity. As such, a practical scheme should take this fact into account to be able to face heuristic embedding schemes which primarily rely on this imperfection of empirical detectors.

V. Results

In this section, we present the results of the different approaches presented in the previous sections. Two effects will be highlighted in this section :

1) The impact of estimating the variance either using (4) (i.e. relying only on the processing pipeline and the photonic noise) or (29) (i.e. also taking into account the image content).
2) The fact that the nature of the dataset will favor different strategies in terms of distribution of the payload in the noise of the DCT coefficient versus in the noise associated to the content.

To do so we used several different datasets with different properties.

a) E1Base: 200 RAW images taken with a E1 Camera at ISO100,

b) CanonBase: 119 RAW images taken with a Canon EOS500D at ISO1600 From ALASKA dataset [22]

c) BOSS dataset without M9 camera: 7642 RAW images taken with 6 different cameras from BOSS dataset [23] excluding the M9 camera whose distribution of the photonic noise is peculiar (see [24], Fig. 2).

The processing pipeline for each dataset are presented in Table IV. We mainly focus on a purely linear pipeline as well as on a pipeline which closely mimics the classical Bossbase pipeline. The focus on the linear pipeline allows us to study our method's performance when the adopted variance estimation model is correct. More precisely, it is one of the few cases in which the processing pipeline can be modeled as a linear transformation which is the core assumption in our model the covariance of DCT coefficients.

On the other hand, the focus on the BOSS pipeline gives a better idea of performances under the standard laboratory settings. In addition, because the linear assumption is certainly the most restrictive, it is important to assess the robustness of our model with respect to the non-linearity of a pipeline

To understand the following experiments, it is important to distinguish between the variance estimation methods and the embedding strategy. To that end, we provide Table III as a summary of our nomenclature for the different embedding schemes used in this paper.

The parameters of the photonic noise c_1 and c_2 were estimated as described in Section II-B. The $\mathbf{H}$ matrix is estimated once for each camera and each processing pipeline using a simple least square regression. To that end, we use a synthetic constant RAW image to which sensor noise is added. This image is then processed using the relevant processing pipeline for each datasets. The RAW and developed images are then reshaped as arrays of 10×10 and 8×8 blocks respectively. We eventually compute $\mathbf{H}$ using Eq. (5). This implies that the covariance matrix of each block was estimated without using neighboring blocks. Even though the estimation for one block should theoretically be carried out with all its neighboring blocks as discussed in Section II, extensive experiments with the E1Base showed no observable gain in security when using those neighboring blocks for the estimation.

Finally, we used the noisy value of the photo-site as an estimate of its mean value for the computation of the variance map instead of denoising the RAW image first. While we also conducted the following experiments using BM3D to denoise the RAW images, the difference in the resulting variance map was always negligible with respect to the one obtained with no denoising and as a consequence no significant security gains were observed.

Due to high number of experiments, the steganalysis was almost always performed using DCTR and the Low-Complexity Linear Classifier [20]. We also performed steganalysis for most experiments with the current state-of-the-art deep neural network EfficientNet-b3 [33] as it was recently shown to reach state-of-the-art performances in the ALASKA2 [32] competition. The stem stride of Efficient-net was set to 1 and training was performed with curriculum beginning from the highest payload used for each dataset with a learning rate of 0.005, divided by 2 on loss plateau. All the relevant code – embedding and variance estimation – will be made available online upon acceptance of the paper.

TABLE III: Nomenclature of the embedding schemes

Prefix	Meaning
Σ-	Uses the noise variance map estimated using the method described in Section II.
c-	Uses the content variance map estimated using the method described in Section IV.
Σ-c-	Uses the sum of the noise and content variance maps.
SI-	Uses the rounding errors as side-information.
Suffix	**Meaning**
Gaussian	Minimizes the power of the MP detector in the continuous domain as described in Section III.
Discrete	Minimizes the power of the MP detector in the quantized domain as described in [3].
UNIWARD	Distortion based schemes as described in [6].
JMiPOD	Uses the variance estimator **and** the embedding strategy described in [3] and compute the variances in the DCT domain using Eq (29) (see [5]).

A. Performance of the Proposed Method over Different Datasets

In this subsection we study the performance of our method on the three aforementioned datasets with different processing

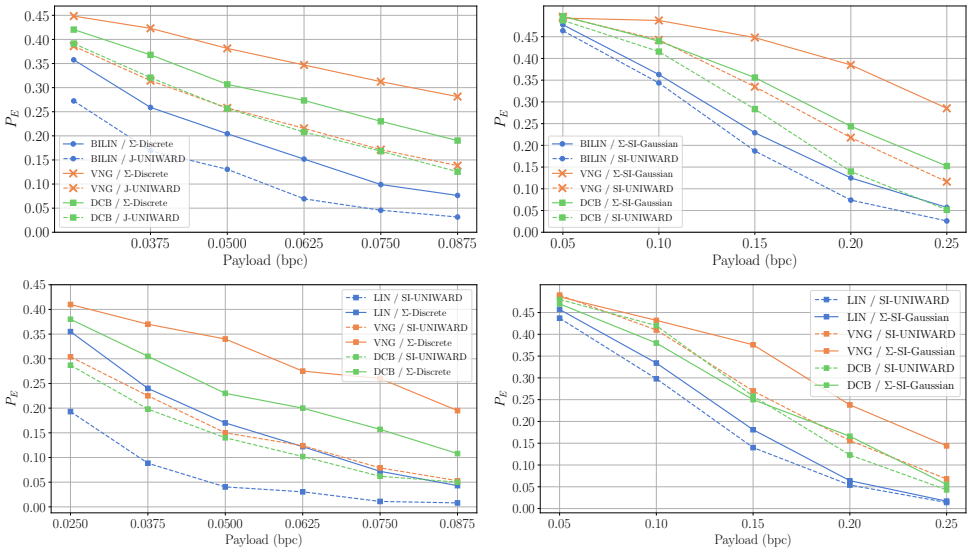


Fig. 7: P_E as a function of payload in bits per coefficients (bpc) for different demosaicking algorithm with the E1Base pipelines. All images are JPEG at QF100. Upper images are steganalyzed with DCTR while lower images were steganalyzed with Efficient-b3.

TABLE IV: Names and operations of the processing pipelines used in the experiments. Gamma correction is never performed except when explicitly stated. The operations are performed in the order they are presented in the table

Pipeline name	**Demosaicking**	**White Balance**	**RGB to grey**	**Downsampling method**	**QF**
E1Base pipelines	Bilinear, VNG or DCB	No	Yes	Crop, 256×256	100
CanonBase pipeline	Bilinear	No	Yes	Resize from 512×512 to 256×256, Bilinear kernel	100
CanonBaseQF pipelines	Bilinear	No	Yes	Crop, 256×256	100,95
Linear Pipeline	Bilinear	No	Yes	Edge crop, 256×256	100
BOSS Pipeline	PPG	Yes, Camera	Yes	Resize from 768×768 (Edge crop) to 256×256, Lanczos kernel,	100

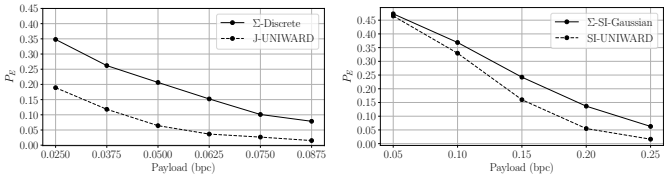


Fig. 8: P_E as a function of payload in bpc for downscaled images on CanonBase with the CanonBase pipelines. All images are JPEG at QF100.

pipelines. E1Base and CanonBase were developed using the E1Base pipelines and CanonBase(QF) pipelines respectively [5] to produce datasets of 5000 images. The BOSS dataset was similarly developed but each image was cropped once in order to maximize content. To do so, we used the edge detector described in [9] and chose the crop containing the maximum number of pixels considered as part of an edge.

Results are presented in Figure 7-9 for E1Base and CanonBase and in Figure 10 for BossBase. Note that VNG and PPG break the stationary assumption since the interpolation of each pixel depends on the value of their gradient with respect to the neighboring photo-sites.

From these results, one can clearly see the impact of the dataset on the performance of our algorithm. Indeed, using our methodology to estimate the variance on E1Base and CanonBase invariably leads to a clear gain of performance by several percents in terms of P_E independently of the processing pipeline. In particular, even in setting with low overall noise, such as when the image is down-sampled or quantized with quantization step $>> 1$, our method still clearly outperforms the UNIWARD family.

However the situation is reversed on the BOSS dataset when using DCTR where Σ-SI Gaussian outperforms SI-UNIWARD only when the BOSS pipeline is used. Interestingly in this case, even for a dataset where all the assumptions of our method are met - in the case of the linear and stationary processing pipeline - the UNIWARD family significantly outperforms Σ-MiPOD and Σ-SI Gaussian. Most interestingly, this situation does not transfer when using EfficientNet. Indeed, in this case, our methods never perform worse than the UNIWARD family and even outperforms it significantly when the knowledge of the rounding errors is used.

In the following, we propose to link this effect to the fact that datasets constructed from BossBase contains more content that is difficult to estimate. As such, hiding only in the noise can be a sub-optimal strategy when facing an empirical detector. This rationale is justified by the fact that E1Base and CanonBase are both made from several crops of the same image, leading to several images with few or smooth content. On the other hand, the BOSS datasets are constructed from only one crop per image, crops which specifically maximize

[5]The processing pipelines use the *rawpy* and *Pillow* libraries

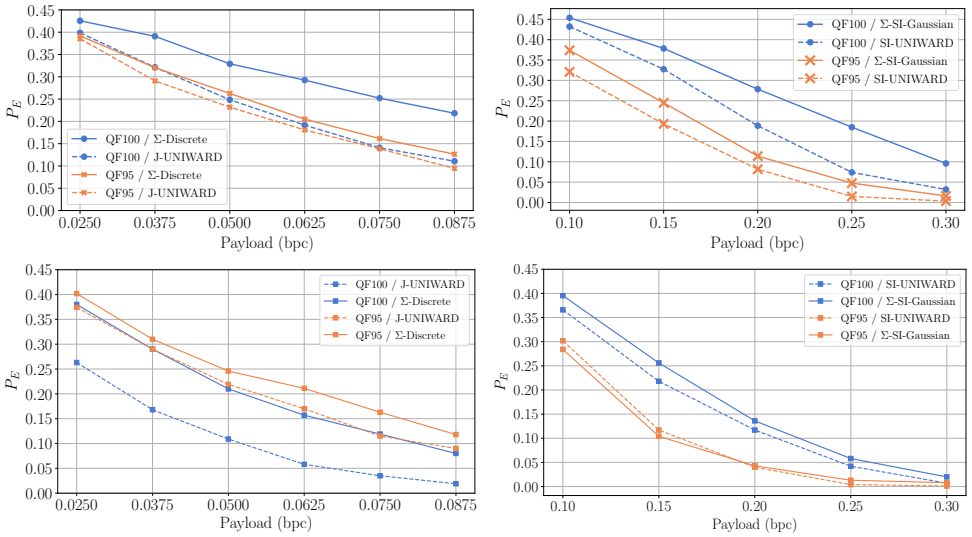


Fig. 9: P_E as a function of payload in bpc for different JPEG quality factors (QF) with CanonBaseQF pipelines. Upper images are steganalyzed with DCTR while lower images were steganalyzed with Efficient-b3.

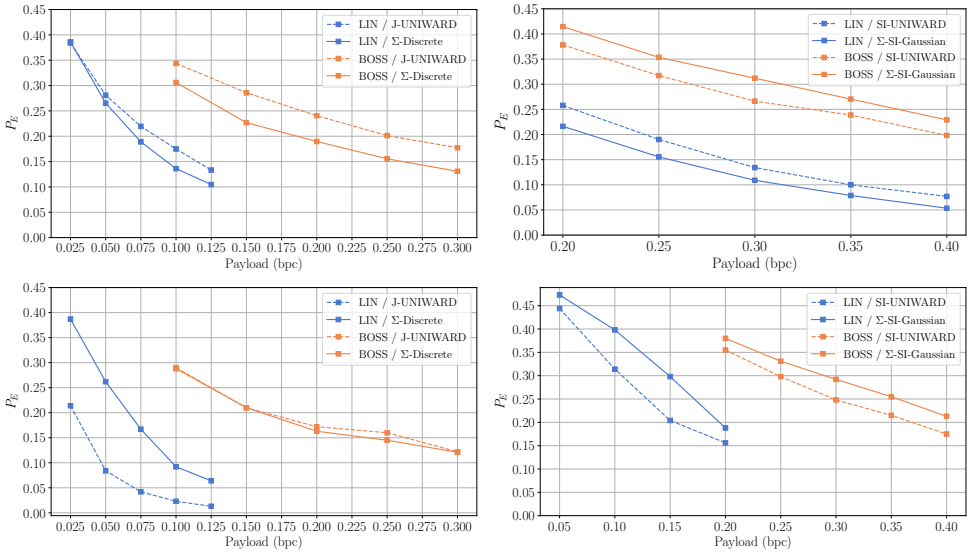


Fig. 10: P_E as a function of payload in bpc for images on Bossbase with two different processing pipeline. The first being the linear processing pipeline, the second being the BOSS pipeline. All images are JPEG at QF100. Steganalysis was performed using DCTR and the LCLC on the upper figures while it was performed with EfficientNet-b3 on the lower ones.

the "difficult" content of the image.

B. Impact of Considering Image Content

In this subsection, we study the effectiveness of different methods in order to take image content into account to improve steganographic security. We thus study the applicability of the methodology presented in Section IV to improve our proposed schemes in these situations. We apply this methodology on the BOSS dataset developed with the linear and BOSS pipelines. Results are presented in Figure 11 and 12.

From these figures, we see that using both the noise and "content" variance always leads to the best results. There is one exception on the BOSS dataset developed with the BOSS pipeline where Σ-c-SI-Gaussian performs the worst (see Fig. 12, right). This can be explained by the fact that both Σ-SI-Gaussian and c-SI-Gaussian both perform better than SI-UNIWARD on this dataset; using the sum of their variance might thus lead to a variance that is so large it becomes insecure.

It is interesting to note that all the studied methods which are based on using the variance due to the content are all more or less on par in terms of P_E. In particular, it is extremely interesting to note that JMiPOD and SI-Gaussian with JMiPOD variance are both highly competitive on these datasets, even beating SI-UNIWARD in the case of the BOSS pipeline. Since the variance estimation of J-MiPOD is extremely crude for the reasons exposed in the introduction, it confirms that the gains in performance observed is not due to a better model

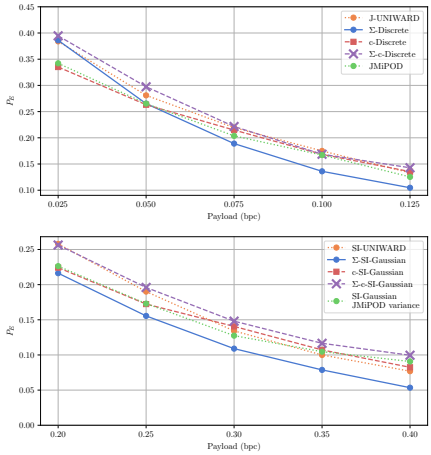


Fig. 11: P_E as a function of payload in bpc for images on Bossbase with the linear processing pipeline comparing the performance of the steganography depending on the variance map used. All images are JPEG at QF100.

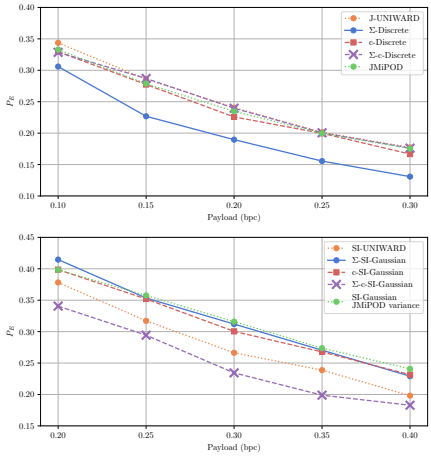


Fig. 12: P_E as a function of payload in bpc for images on Bossbase with the BOSS pipeline comparing the performance of the steganography depending on the variance map used. All images are JPEG at QF100.

but to the fact that these schemes capitalize on the difficulty of the steganalyst to estimate some part of the content. This phenomenon was already mentioned in [3].

C. Effect of Source and Content Diversity

In this subsection we investigate the impact of content diversity in a dataset on the performance of steganography. In particular, we highlight the fact that the performance of the UNIWARD family is mostly due to the failure of the steganalyst to estimate the model parameters for each image.

To do so, we produce a dataset where this estimation is almost assured to be perfect. Using the observation in Section IV that empirical detectors can capitalize on the inter-dependencies between images of the dataset to reduce the impact of bad content estimation, we took one RAW image

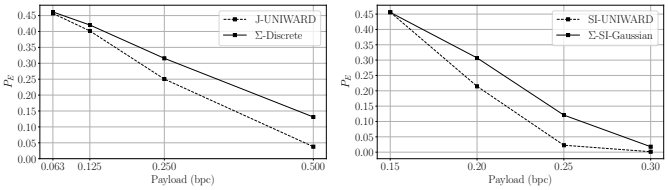


Fig. 13: P_E as a function of the payload in bpc on 5000 noisified images from one BossBase image with the linear processing pipeline. All images are JPEG at QF100.

from BossBase and simulated 5000 noisified version of this RAW image by adding a simple heteroscedastic noise to the photo site with $c_1 = 1$ and $c_2 = 0$. These images were then developed using the linear processing pipeline and cropping to 512×512 the area given by the edge crop strategy on the "noiseless" image. In this situation, the diversity of the content and of the source is reduced to its minimum since only the photonic noise values differ between images and not the content. We compare the performance of J-UNIWARD, SI-UNIWARD, $\Sigma-$Discrete and Σ-SI-Gaussian on this dataset in Figure 13. These results show a clear under-performance of the UNIWARD family with respect to Σ schemes, similar to what is observed on the E1Base and CanonBase giving credit to the aforementioned hypothesis.

D. Validation of the Variance Estimation Method

In this subsection we study the effectiveness of the variance estimation method described in Section II in improving steganographic security.

Because of the independence assumption, one might wonder if there is a real gain in security by keeping the dependencies between pixels and DCT coefficients up to JPEG compression for the variance estimation. We here repeat the experiment in Section V-A on E1Base with a purely linear processing pipeline and the pipeline using the DCB demosaicking algorithm. However, we now compare the performance of $\Sigma-$Discrete to the performance of Discrete which used a variance map estimated differently. Instead of keeping the full covariance matrix up until the end in the estimation process, as explained in Section II, we drop all covariances after the RGB to grey conversion, keeping only a diagonal covariance. The variances are then computed in the DCT domain using Eq. (29). This amounts to assuming pixels are independent after the RGB to grey conversion. Results are presented in Figure 14. We also validate our approach for taking into account gamma correction on E1Base with the linear processing pipeline with gamma correction activated with rawpy default parameters $(2.222, 4.5)$ in Figure 15.

In Figure 14 we clearly observe that dropping the dependencies in the spatial domain during the variance estimation process leads to extremely poor security performance. Indeed, keeping the dependencies until the end of the process leads to a gain of 15% on average in terms of P_E. This clearly outlines the importance of taking into account the dependencies created by the processing pipeline as well as validate our estimation method.

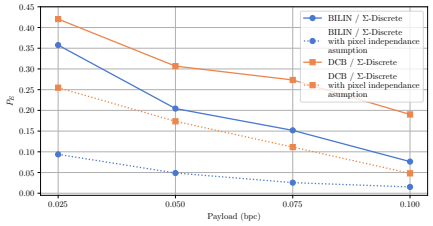


Fig. 14: P_E as a function of payload in bits per coefficients (bpc) on E1Base processed with E1Base pipelines. It compares the variance estimation method when independence is assumed in the spatial domain or in the DCT domain. All images are JPEG at QF100.

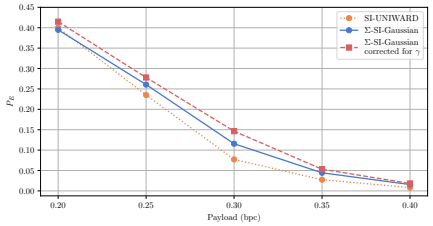


Fig. 15: P_E as a function of payload in bits per coefficients (bpc) on E1Base processed with E1Base pipeline with bilinear demosaicking and gamma correction. All images are JPEG at QF100.

VI. Discussion and Analysis

A. Dataset difficulty

In the preceding section, we conjectured that the loss of performance observed with our method between E1Base and BOSSBase was due to the fact that the variance due to error in the content estimation dominated the variance due to the sensor noise. Our method only takes sensor noise into account in its model, contrary to the UNIWARD family which was specifically designed to embed in location where content is difficult to estimate.

To motivate this conjecture we propose to use the variance due to the modeling error as described in Section IV. As was observed in Figure 3, this variance is a good indicator of content as it increases in textured areas and at sharp transitions. We then compute the average variance for each image of each dataset developed with the Linear pipeline as it is the pipeline where our embedding scheme under-performs when using rich models. We finally plot the histograms of average variance in Figure 16.

As can be observed from these histograms, the average "content" variance for E1Base is close to zero. The situation is quite different for BOSSBase, where the "content" variance has a far greater range up to 0.5.

B. Impact of the quantization on the noise covariance

Most experiments in this paper were done using relatively high quality factors. In general, the steganographer should avoid cover with low quality factors as stronger quantization leads to lower variance of the DCT coefficient and consequently, *mutatis mutandi*, lower security [31].

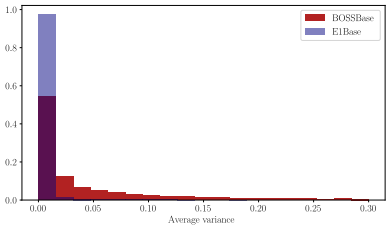


Fig. 16: Histogram of the average variance due to error in modeling content – aka "content" variance – per image for BOSSBase and E1Base.

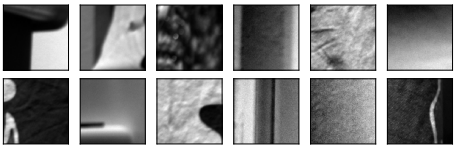


Fig. 17: Examples of images taken from the E1Base dataset processed with the linear pipeline.

However, in our case, the reason not to use low QF is directly linked to the variance estimation method we outlined in Section II. Indeed, not only does quantization lower the variances of DCT coefficient but also their covariances. To illustrate this phenomenon, we give in Figure 19 a series of a covariance matrix of the same image block as the quality factor decreases. Note that these covariances are obtained **before** the rounding of DCT coefficients. One can observe that most of of the covariances become negligible as soon as quantization with QF95 is performed. Furthermore, even though the structure of the covariance matrices does not change much for even lower QFs, the variances and covariances tend to zero quickly as the QF decreases.

If we also take into account the rounding of the DCT coefficient, the impact on the structure is even more dramatic as most of higher modes will actually have a variance far smaller than the quantization step for lower quality factors. Note however that is effect is highly dependent on the power of the sensor noise.

VII. Conclusion

In this paper, we propose a new method to estimate covariances and variances in the JPEG domain using the knowledge of the processing pipeline and a multivariate Gaussian model of the DCT coefficients noise. By leveraging dependencies between pixels and DCT coefficients, we are able to get precise

Fig. 18: Examples of images taken from the BOSSBase dataset processed with the linear pipeline.

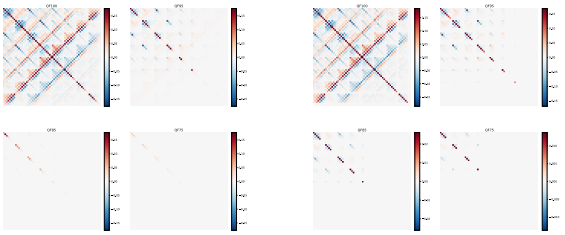


Fig. 19: Covariance matrix of a block of DCT coefficient of image *1200* in BOSSBase with the linear pipeline for different quality factors. On the left are covariances matrices obtained before rounding and on the right after rounding.

estimates of the variances in the DCT domain. The method is also shown to be robust to pipelines diverging from its linearity and stationary assumptions. We also propose a novel, side informed, model-based scheme in the JPEG domain based on minimizing the power of the most powerful detector in the continuous domain while setting the message size constraint in the discrete domain. An exact, analytical, expression of the power of the MP detector is also derived. Finally, since the proposed model only deals with the DCT coefficients noise, a strategy is presented in order to take into account content complexity through an estimation of the variance of the estimation error of the content. Evidently, this scheme, which is based on a likelihood ratio test, does not bring the same security level as other adversarial schemes based on deep neural nets because of the model mismatch between the LRT and the true generative process of the image. However, it does not require computationally expensive tasks such as the generation of multiple datasets of stego contents nor the training of deep neural networks.

These three contributions lead also to several insights. Firstly, the trade-off between noise and content-complexity in steganography is highlighted, and this steganographic scheme greatly outperforms the state-of-the-art for settings where noise predominates over content. Secondly, to maintain the performance of the scheme for datasets with complex contents w.r.t distortion based schemes, we show the importance to estimate the noise related to the content estimation error.

A last conclusion is related to the importance of dependencies for steganalysis since an important gap in performance is highlighted between detectors taking into account the full covariance of DCT blocks and detectors not considering correlations between pixels.

Future works will thus include the extension of our scheme to non-additive schemes able to use the full covariance matrix. On a more practical matter, we will study the feasibility of estimating the covariance matrix without the knowledge of the RAW image.

ACKNOWLEDGMENTS

This work has been funded in part by the French National Research Agency (ANR-18-ASTR-0009), ALASKA project: https:// alaska.utt.fr, by the French ANR DEFALS program (ANR-16- DEFA-0003).

APPENDIX
ASYMPTOTIC PERFORMANCE OF THE LRT

In this section we derive the exact asymptotic performance of test proposed in Section III-B. Let :

$$p_{\sigma_i}(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-x^2}{2\sigma_i^2}\right), \tag{30}$$

$$q_{\sigma_i,\epsilon_i}(x) = \frac{1}{\sqrt{2\pi(\sigma_i^2+\epsilon_i^2)}} \exp\left(\frac{-x^2}{2(\sigma_i^2+\epsilon_i^2)}\right), \tag{31}$$

$$\Lambda_i(z,\sigma_i,\epsilon_i) = \ln\left(\frac{p_{\sigma_i}(z)}{q_{\sigma_i,\epsilon_i}(z)}\right), \tag{32}$$

$$\Lambda(\mathbf{z},\sigma,\epsilon) = \sum_{i=0}^{N} \Lambda_i(z_i,\sigma_i,\epsilon_i), \tag{33}$$

We compute the first two moments of the LRT under each hypothesis. We begin by linking the expectation to the KL-divergence between two Gaussian :

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_0}[\Lambda_i] &= \int_{-\infty}^{\infty} \ln\left(\frac{q_{\sigma_i,\epsilon_i}(z)}{p_{\sigma_i}(z)}\right) p_{\sigma_i}(z)\,dz \\ &= -D_{KL}(p||q) \\ &= \ln\left(\frac{\sqrt{\sigma_i^2+\epsilon_i^2}}{\sigma_i}\right) + \frac{\sigma_i^2}{2(\sigma_i^2+\epsilon_i^2)} - 0.5.\end{aligned} \tag{34}$$

Using the same argument for the expectation under $\mathcal{H}_1$ we obtain:

$$\mathbb{E}_{\mathcal{H}_1}[\Lambda_i] = D_{KL}(q_{\sigma_i,\epsilon_i}||p_{\sigma_i}) \tag{35}$$

The variances can also be computed analytically by identifying the moments of the Gaussian in the integral:

$$\begin{aligned}Var_{\mathcal{H}_0}[\Lambda_i] = &\int_{-\infty}^{\infty} \ln^2\left(\frac{q_{\sigma_i,\epsilon_i}(z)}{p_{\sigma_i}(z)}\right) p_{\sigma_i}(z)\,dz \\ &- D_{KL}^2(p_{\sigma_i}||q_{\sigma_i,\epsilon_i}).\end{aligned} \tag{36}$$

Let:

$$c_1 = \ln^2\left(\frac{\sigma_i}{\sigma_i^s}\right), \quad c_2 = 2\ln\left(\frac{\sigma_i^s}{\sigma_i}\right)\frac{(\sigma_i^s)^2-\sigma_i^2}{2(\sigma_i^s)^2\sigma_i^2},$$
$$c_3 = \left(\frac{(\sigma_i^s)^2-\sigma_i^2}{2(\sigma_i^s)^2\sigma_i^2}\right)^2,$$

with $(\sigma_i^s)^2 = \sigma_i^2+\epsilon_i^2$.
We can rewrite the variance as:

$$\begin{aligned}Var_{\mathcal{H}_0}[\Lambda_i] = &\int_{-\infty}^{\infty} \left(c_3 z^4 + c_2 z^2 + c_1\right) p_{\sigma_i}(z)\,dz \\ &- D_{KL}^2(p_{\sigma_i}||q_{\sigma_i,\epsilon_i}).\end{aligned} \tag{37}$$

Finally, recognizing the second and fourth moment of the Gaussian distribution, we obtain:

$$Var_{\mathcal{H}_0}[\Lambda_i] = 3c_3\sigma_i^4 + c_2\sigma_i^2 + c_1 - D_{KL}^2(p_{\sigma_i}||q_{\sigma_i,\epsilon_i}), \tag{38}$$

and similarly under $\mathcal{H}_1$.

With some routine calculations, the second moments of the LRT can be simplified as:

$$\begin{cases} Var_{\mathcal{H}_0}\left[\Lambda_i\right] = \frac{\epsilon_i^4}{2\left(\epsilon_i^2+\sigma_i^2\right)^2} \\ Var_{\mathcal{H}_1}\left[\Lambda_i\right] = \frac{\epsilon_i^4}{2\sigma_i^4} \end{cases} \tag{39}$$

Finally, using the independence of DCT coefficients, the full moments are obtained as the sum of the individual moments:

	$\mathbb{E}\left[\Lambda\right]$	$Var\left[\Lambda\right]$
$\mathcal{H}_0$	$-\sum_{i=1}^{N} D_{KL}\left(p_{\sigma_i}\|\|q_{\sigma_i,\epsilon_i}\right)$	$\sum_{i=1}^{N} \frac{\epsilon_i^4}{2(\epsilon_i^2+\sigma_i^2)^2}$
$\mathcal{H}_1$	$\sum_{i=1}^{N} D_{KL}\left(q_{\sigma_i,\epsilon_i}\|\|p_{\sigma_i}\right)$	$\sum_{i=1}^{N} \frac{\epsilon_i^4}{2\sigma_i^4}$

As the number of DCT coefficient $N \to \infty$, Linderberg's central limit theorem implies that:

$$\Lambda(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\epsilon}) \rightsquigarrow \begin{cases} \mathcal{N}\left(\mathbb{E}_{\mathcal{H}_0}\left[\Lambda\right], \mathrm{Var}_{\mathcal{H}_0}\left[\Lambda\right]\right), & \text{under } \mathcal{H}_0 \\ \mathcal{N}\left(\mathbb{E}_{\mathcal{H}_1}\left[\Lambda\right], \mathrm{Var}_{\mathcal{H}_1}\left[\Lambda\right]\right), & \text{under } \mathcal{H}_1 \end{cases} \tag{40}$$

where $\rightsquigarrow$ denotes convergence in distribution. From the limiting distribution of the LR, the asymptotic power of the LRT is thus given by:

$$P_D = \mathbb{P}\left(\delta\left(x\right) = \mathcal{H}_1 | \mathcal{H}_1\right) \tag{41}$$

$$= Q\left(\frac{Q^{-1}\left(P_{FA}\right)\sqrt{\mathrm{Var}_{\mathcal{H}_0}\left[\Lambda\right]} + \mathbb{E}_{\mathcal{H}_0}\left[\Lambda\right] - \mathbb{E}_{\mathcal{H}_1}\left[\Lambda\right]}{\sqrt{\mathrm{Var}_{\mathcal{H}_1}\left[\Lambda\right]}}\right), \tag{42}$$

where Q is the tail distribution function of the standard normal distribution.

REFERENCES

[1] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 920–935, sep 2011.

[2] V. Sedighi, J. J. Fridrich, and R. Cogranne, "Toss that bossbase, Alice!," in *Media Watermarking, Security, and Forensics*, Feb 2016, Proc. IS&T.

[3] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, Feb 2016.

[4] M. Sharifzadeh, M. Aloraini, and D. Schonfeld, "Quantized gaussian embedding steganography," in *ICASSP*, May 2019, IEEE.

[5] T. Denemark and J. Fridrich, "Model based steganography with precover," *Electronic Imaging*, vol. 2017, no. 7, pp. 56–66, Jan 2017.

[6] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, jan 2014.

[7] K. Chen, H. Zhou, W. Zhang and N. Yu " Defining Cost Functions for Adaptive JPEG Steganography at the Microscale" *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 1052–1066, Apr 2019.

[8] Q. Giboulot, R. Cogranne, and P. Bas, "JPEG steganography with side information from the processing pipeline," in *ICASSP*, IEEE, Barcelone, Spain, May 2020.

[9] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, oct 2008.

[10] A. Foi, "Practical denoising of clipped or overexposed noisy images", Proc. 16th European Signal Process. Conf., EUSIPCO 2008, pp. 1-5 August 2008.

[11] T.H. Thai, R. Cogranne, and F. Retraint, "Camera model identification based on the heteroscedastic noise model," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 250–263, Jan 2014.

[12] T.H. Thai, R. Cogranne, and F. Retraint, "Statistical model of quantized DCT coefficients: Application in the steganalysis of jsteg algorithm," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1980–1993, May 2014.

[13] A. Foi, "Practical denoising of clipped or overexposed noisy images," in *2008 16th European Signal Processing Conference*. IEEE, 2008.

[14] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.

[15] T. Taburet, P. Bas, J. Fridrich, and W. Sawaya, "Computing dependencies between DCT coefficients for natural steganography in JPEG domain," in - *IHMMSec2019*.

[16] E.L. Lehmann and J.P. Romano, *Testing Statistical Hypotheses, Second Edition*, Springer, 3rd edition, 2005.

[17] D Dowson and A Wragg, "Maximum-entropy distributions having prescribed first and second moments (corresp.)," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 689–693, 1973.

[18] J. Butora and J. Fridrich, "Minimum perturbation cost modulation for side-informed steganography," in *Electronic Imaging*, *Proc. IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics* San Francisco, United States, Jan. 2020.

[19] V. Holub and J. Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, Feb 2015.

[20] R. Cogranne, V. Sedighi, J. Fridrich, and T. Pevny, "Is ensemble classifier needed for steganalysis in high-dimensional feature spaces?," in *International Workshop on Information Forensics and Security (WIFS)*. Nov 2015, IEEE.

[21] M. Kivanc Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *ICASSP*, 1999, IEEE.

[22] R. Cogranne, Q. Giboulot, and P. Bas, "The alaska steganalysis challenge: A first step towards steganalysis," in *IHMMSec'19*, Paris, France, July 2019, pp. 125–137, ACM.

[23] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system — the ins and outs of organizing BOSS," in *Information Hiding, 13th International Workshop*, Prague, Czech Republic, May 18–20, 2011, pp. 59–70, LNCS vol.6958.

[24] T. Denemark, P. Bas, and J. Fridrich, "Natural Steganography in JPEG Compressed Images," in *Electronic Imaging*, *Proc. IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics* San Francisco, United States, Jan.2018.

[25] J. Kodovsky and J. Fridrich, "Steganalysis of JPEG Images Using Rich Models," in *Electronic Imaging*, *Proc. IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics* San Francisco, United States, Jan.2012.

[26] S. Xiaofeng, L. Fenlin, Y. Chunfang, Luo. Xiangyang and Z. Yi, "Steganalysis of adaptive JPEG steganography using 2D Gabor filters," in *IH&MMSEC 2015*, 2015

[27] M. Boroumand, M. Chen, J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2018.

[28] R. Cogranne, Q. Giboulot, and P. Bas, 'Steganography by Minimizing Statistical Detectability: The cases of JPEG and Color Images," in *IHMMSEC 2020*, Denver, USA, Jun 2020

[29] W. Tang, B. Li, S. Tan, M. Barni, and J. Huang. Cnn-based adversarial embedding for image steganography. *IEEE Transactions on Information Forensics and Security*, 2019.

[30] S. Bernard, P. Bas, J. Klein, and T. Pevný. Explicit Optimization of min max Steganographic Game. *IEEE Transactions on Information Forensics and Security*, 2020.

[31] J. Butora, J. Fridrich: Effect of JPEG Quality on Steganographic Security. in *IHMMSec'19*, Paris, France, July 2019, pp. 47–56, ACM.

[32] R. Cogranne, Q. Giboulot, P. Bas. ALASKA-2 : Challenging Academic Research on Steganalysis with Realistic Images *to be published* in *International Workshop on Information Forensics and Security (WIFS)*. Nov 2020, IEEE.

[33] T. Mingxing, L. Quoc V. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946, 2019