Are we there yet?

Mehdi Boroumand, Rémi Cogranne, Jessica Fridrich

Are we there yet?

Mehdi Boroumand,[+] *Remi Cogranne,*[×] *and Jessica Fridrich,*[+]
[+] *Department of ECE, SUNY Binghamton, NY, USA, {mboroum1,fridrich}@binghamton.edu*
[×] *Troyes University of Technology (UTT), Lab. of System Modelling and Dependability (LM2S), France, Remi.Cogranne@utt.fr*

Abstract

The purpose of this study is to prepare a source of realistic looking images in which optimal steganalysis is possible by enforcing a known statistical model on image pixels to assess the efficiency of detectors implemented using machine learning. Our goal is to answer the questions that researchers keep asking: "Are our empirical detectors close to what can be possibly detected? How much room there is for improvement?" or simply "Are we there yet?" Our goal is achieved by applying denoising to each image in a dataset of real images to remove complex statistical dependencies introduced by processing and, subsequently, adding noise of simpler and known statistical properties that allows deriving a closed form expression of a likelihood ratio test. The theoretical upper bound informs us about the amount of further possible improvement. Three content-adaptive stego algorithms in the spatial domain and simple LSB matching are used to assess the performance of a convolutional neural network detector and a detector based on rich models with respect to the derived upper bound on performance. The short answer to the posed question is "We are much closer now but there is still non-negligible room for improvement."

Motivation

Steganography is the art of covert communication in which messages are hidden in cover objects so that the very existence of the secrets cannot be established. The objective of steganalysis is to detect the usage of steganography and do so as reliably as possible. A popular choice for cover objects today are digital multi media files, such as digital images, audio, and video. Such objects are ideal for covert communication for two reasons. They contain an indeterministic component, the acquisition noise, that helps mask the presence of steganographic embedding changes. Additionally, the inherent complexity of these objects is hard to capture using tractable and estimable statistical models, which further complicates detection. Steganographers fine-tune their embedding algorithms to locally adapt to content complexity since complicated textures and small-scale details are extraordinarily difficult to model statistically. This forced steganalysts to use complex high-dimensional (rich) media models [14, 20, 8, 15, 16, 25, 10, 11, 2] and, recently, non-linear hierarchical models with a large number of parameters, deep neural networks [23, 1, 29, 28, 30, 31, 34, 6, 27, 33].

It should be stressed that, fundamentally, it is the unavailability of statistical models for natural images that is responsible for this seemingly never ending spiral development. Steganography in artificial sources (sources with a known statistical model) can be perfectly secure[1] as covers can be synthesized [3] to communicate at a positive rate (payload whose size is linear w.r.t. the number of cover elements) [22, 26]. Likewise, optimal detectors of imperfect steganography methods in artificial sources can be constructed and their performance computed.

The situation is quite different for empirical sources that lack description using tractable and estimable statistical models. All steganographic methods inevitably become imperfect and the size of their secure payload sublinear in the number of pixels due to the so-called square root law [19, 13, 18, 17]. Detectors can be built that can distinguish between cover and stego objects better than randomly guessing. Without a cover model, however, we are unable to assess how good our steganography methods are and how well our detectors perform.

This paper is an attempt to address this problem by forming an artificial source of realisticly looking images while forcing a known statistical model on pixels to allow derivation of optimal statistical tests for benchmarking empirical detectors built using machine learning. While it is entirely possible to synthesize artificial images for this purpose, the authors believe that it is valuable to keep a more realistic dataset with images visually similar to popular sources, such as BOSSbase 1.01 [4], in which content adaptive schemes execute changes with a similar selection channel as in the original source. We also need to avoid sources in which steganography would be too easy or too hard to detect while making sure that an optimal detector can be derived. Since these requirements are in conflict, preparing a suitable source of both cover and stego images is quite challenging

The idea for the cover source proposed in this paper was inspired by the experiment reported in Fig. 5 of [24]. The authors selected one BOSSbase image, denoised it, and then created 10,000 different versions of the same image by adding to it 10,000 independent realizations of sensor acquisition noise. Steganalysis in such a homogeneous cover source with the spatial rich model (SRM) [14] and MiPOD embedding algorithm [24] was reported to be rather close in terms of the Receiver Operating Characteristic (ROC) to the optimal statistical test designed for the noise component. However, for a *heterogeneous* source with images of diverse content, the SRM detector lagged behind the optimal Likelihood Ratio Test (LRT) quite a bit most likely due to the inability of the empirical detector to deal with

[1] In Cachin's sense [7].

the diversity of natural images (Fig. 6 in [24]).

The strategy adopted in this paper is to start with an existing dataset, apply a denoising filter to all images to remove complex noise introduced during acquisition and the subsequent development of the image from the raw sensor capture to a viewable form. Then, independent realizations of a Gaussian noise whose variance was estimated per pixel from the original images is reintroduced to force a known and tractable noise model in the cover source. This needs to be executed with care to prevent introducing dependencies among stego pixels. In particular, the pixels costs cannot be computed from the cover itself as the stego pixels would be dependent, which would prevent derivation of a closed-form LRT.

The proposed cover source dataset is described in the next section. TBD

Cover source

In this section, we first describe in detail the cover source preparation and then discuss the specific choices that were made.

The cover source was generated the from the union of BOSSbase 1.01 [4] and BOWS2 [5] grayscale images resized from their original 512×512 size in Matlab (using '`imresize.m`' with default parameters) to $20,000$ 256×256 grayscale images. The smaller image size was selected in anticipation that the best empirical detectors will be deep neural networks, which typically require smaller images for effective training to fit reasonable size minibatches to the memory of current GPUs. We note that the leading performers in the spatial domain, the YeNet [32], YedroudjNet [33], and the SRNet [6], were trained and benchmarked on this same database. The union of both databases with $20,000$ 256×256 grayscale images will be denoted $\mathcal{B}$. We note that should future deep architectures require more than $20,000$ images, cropping instead of resizing the original images would produce four times as many, $80,000$ images, for a bigger dataset.

The formation of the dataset is explained in five steps and an additional, sixth, step needed for creating the stego images in a way that allows computation of an optimal statistical test.

Step 1: Estimate pixel variance

MiPOD's variance estimator (Section V in [24]) was applied to all 20,000 images from $\mathcal{B}$ to estimate the local variance of pixels' noise residual. For a given image and its pixel (i,j), $1 \le i,j \le 256$, we denote its estimated variance σ_{ij}^2. Note that the output of MiPOD's estimator is lower bounded: $\sigma_{ij}^2 \ge 0.01$ for all (i,j).

Step 2: Denoising

All images from $\mathcal{B}$ were first denoised to remove complex dependencies among pixels introduced by the RAW developer and subsequent processing. We used the wavelet denoising method described in [21] with Daubechies 8-tap wavelets and standard deviation of the removed Gaussian i.i.d. noise $\sigma_{\text{den}} = 10$. The pixel values in the denoised image were left in their non-rounded form but were clipped to the interval corresponding to 8-bit grayscale images $[0,255]$.

Step 3: Narrowing dynamic range

As the third step, the dynamic range of each denoised and clipped image was narrowed to the range $[15,240]$ by linearly mapping the interval $[0,255]$ to $[15,240]$ using :

$$g(x) = 15 + \frac{225}{255}x. \tag{1}$$

The scaled values were also rounded to integers, which we will denote $\mu_{ij} \in \{15,\dots,240\}$. The resulting 8-bit grayscale image with a narrower dynamic range, which we denote μ_{ij}, will next be noisified with the variances estimated in Step 1 and further adjusted in Step 4.

Step 4: Adjusting the variance

The estimated pixel variances σ_{ij}^2 were adjusted so that the probability of a pixel getting out of the 8-bit dynamic range $[0,255]$ after noisification is at most 2.87×10^{-7}, the probability of a one-sided 5σ-outlier. This was done by making sure that σ_{ij} is smaller or equal to one fifth of the distance between the pixel mean μ_{ij} to the dynamic range boundary (0 or 255) :

$$\sigma'_{ij} = \min\Big\{\frac{1}{5}\min\{\mu_{ij}, 255-\mu_{ij}\}, \sigma_{ij}\Big\}. \tag{2}$$

Additionally, we also introduced a floor for the variance to be able to simplify the optimal test using the fine quantization limit approximation :

$$\underline{\sigma}_{ij} = \max\{\sigma'_{ij}, \sigma_0\}. \tag{3}$$

In this paper, experimented with $\sigma_0 \in \{0.5, 1\}$.

Step 5: Noisifying

The noisified pixel c_{ij} is obtained by adding to μ_{ij} a sample ξ_{ij} from $\mathcal{N}(0, \underline{\sigma}_{ij}^2)$, rounding to an integer and clipping to $[1,254]$ to make sure the embedding will be free to modify all pixels by ± 1 without getting out of the dynamic range. In other words, in our dataset we impose the cover image model from MiPOD – pixels are realizations of independent Gaussian variables $\mathcal{N}(\mu_{ij}, \underline{\sigma}_{ij}^2)$ that are rounded to integers (denoted with the square bracket $[\cdot]$), and then clipped to a finite dynamic range. Symbolically,

$$c_{ij} = [\mu_{ij} + \xi_{ij}] \tag{4}$$

$$c_{ij} = \begin{cases} c_{ij} & \text{if } 1 \le c_{ij} \le 254 \\ 1 & \text{if } c_{ij} \le 0 \\ 254 & \text{if } c_{ij} \ge 255. \end{cases} \tag{5}$$

The cover image pixels thus follow a p.m.f. p_{ij}, $c_{ij} \sim p_{ij}$:

$$p_{ij}(m) = \begin{cases} 0 & m = 0 \\ Q_{ij}\left(m - \frac{1}{2}\right) & m = 254 \\ Q_{ij}\left(m - \frac{1}{2}\right) - Q_{ij}\left(m + \frac{1}{2}\right) & 1 < m < 254 \\ 1 - Q_{ij}\left(m + \frac{1}{2}\right) & m = 1 \\ 0 & m = 255 \end{cases}$$

(6)

with $Q_{ij}(x)$ defined as the tail probability of $\mathcal{N}(\mu_{ij},\underline{\sigma}_{ij}^2)$:

$$Q_{ij}(x) \triangleq \Pr\{\mathcal{N}(\mu_{ij},\underline{\sigma}_{ij}^2) > x\}. \tag{7}$$

The values c_{ij} form the data source used in our experiments. This cover source will be denoted $\mathcal{B}(\sigma_0)$.

Stego images

The stego methods used in this study are all ternary embedding algorithms that change the (i,j)th pixel by ± 1 with equal probability β_{ij}. Since we curbed the cover values in Step 5 to the interval $[1,254]$, the embedding does not need to be constrained in any way – all pixels can be changed either way. For content adaptive steganography, the probabilities (change rates) β_{ij} are determined from pixel costs, which are typically computed from a local neighborhood of pixel (i,j). This dependence is quite complicated as the costs are usually computed in a non-linear fashion from outputs of several high-pass filters. Thus, computing β_{ij} from the noisified cover c_{ij} would create complex dependencies among stego pixels, preventing thus a closed-form expression for the distribution of stego pixels and tractable evaluation of the associated likelihood ratio test.

We resolved this problem by computing the pixel costs (change rates β_{ij}) from the corresponding original image from $\mathcal{B}$. Another possibility is to compute the costs from a different independent noisification of the image. Both versions gave similar results in our experiments. The former was used as default for the rest of this paper.

Since β_{ij} now does not depend on the specific noisification of the image and since the embedding changes are executed independently, the stego pixel p.m.f. is factorizable. In particular, it is a product of the following Gaussian mixtures q_{ij} over all pixels :

$$\begin{aligned} q_{ij}(m) &= (1-2\beta_{ij})p_{ij}(m) + \beta_{ij}p_{ij}(m-1) \\ &\quad + \beta_{ij}p_{ij}(m+1). \end{aligned} \tag{8}$$

For S-UNIWARD, HILL, and WOW, the change rates were obtained from an embedding simulator (e.g., assuming optimal source coding).

Optimal test

Given an image with pixels s_{ij}, the steganalyst is facing the following statistical hypothesis test for all (i,j) :

$$\begin{aligned} &\mathcal{H}_0 : s_{ij} \sim p_{ij} \\ &\mathcal{H}_1 : s_{ij} \sim q_{ij}. \end{aligned} \tag{9}$$

We will assume that the parameters of the added MVG noise, the mean μ_{ij}, and the variance $\underline{\sigma}_{ij}^2$, are known. We also assume that the change rates β_{ij} are known. Under these assumptions, the test is simple, and the optimal statistic is the log-likelihood ratio

$$\Lambda(\mathbf{s}) = \sum_{i,j} \Lambda_{ij} = \sum_{i,j} \log\left(\frac{q_{ij}(s_{ij})}{p_{ij}(s_{ij})}\right) \tag{10}$$

by the statistical independence of pixels. For convenience, we will use the following normalized form of the log-LRT :

$$\Lambda^\star(\mathbf{s}) = \frac{\sum_{i,j}\Lambda_{ij} - E_{\mathcal{H}_0}[\Lambda_{ij}]}{\sqrt{\sum_{i,j} Var_{\mathcal{H}_0}[\Lambda_{ij}]}}, \tag{11}$$

where

$$E_{\mathcal{H}_0}[\Lambda_{ij}] = \sum_{i,j} p_{ij}\Lambda_{ij} \tag{12}$$

$$Var_{\mathcal{H}_0}[\Lambda_{ij}] = \sum_{i,j} p_{ij}\Lambda_{ij}^2 - \left(E_{\mathcal{H}_0}[\Lambda_{ij}]\right)^2. \tag{13}$$

Under the fine quantization limit, $\sigma_0 \le \sigma_{ij}$ for all i,j, and as the number of pixels approaches infinity, the Lindeberg's version of the Central Limit Theorem implies

$$\Lambda^\star(\mathbf{s}) \rightsquigarrow \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\varrho,1) & \text{under } \mathcal{H}_1 \end{cases}, \tag{14}$$

where $\rightsquigarrow$ means convergence in distribution and $\varrho = \sum_{i,j}\underline{\sigma}_{ij}^{-4}\beta_{ij}^2 > 0$ is the deflection coefficient.

Discussion

The denoising step in the preparation of the cover source is essential because we give the means of the MVG (the denoised signal) to the optimal test while the network would have to deal with the complexity of numerous noise sources that occur in natural images. By adding the MVG to the denoised image, we force the cover complexity to be primarily in the noise component. This way, both the optimal test and the network need to deal with the content complexity due to indeterminism, the added noise. Of course, the network still needs to learn a model for the (heterogeneous) denoised content.

Because the added noise mimics the content complexity of the original image (a combination of the indeterminism in the original image and texture), when the change rates β_{ij} are computed from the noisified cover c_{ij}, the costs profile (change rates) of the tested content-adaptive algorithms looks similar to the costs (change rates) computed from the original image from $\mathcal{B}$ (see Figure 1). This justifies our choice of computing the change rates from the original image.

Note that since $p_{ij}(0) = p_{ij}(255) = 0$, the boundary values do not occeur in covers from $\mathcal{B}(\sigma_0)$, and whenever the embedding produces these "forbidden" values, the LRT becomes infinity, arranging for a perfect detection in this case. Fortunately, due to our choice of the standard deviation $\underline{\sigma}_{ij}$ (2)–(3), this occurs with very small probability. In our tests, we typically saw at most one such image.

The scaling and clipping of the dynamic range in Step 3 was also necessary because otherwise the embedding would have to be adjusted not to change the boundary value 255 up and 0 down, which would reintroduce a dependence between a specific noisification c_{ij} and the change rates β_{ij}, preventing the derivation of a closed-form expression for the LRT.

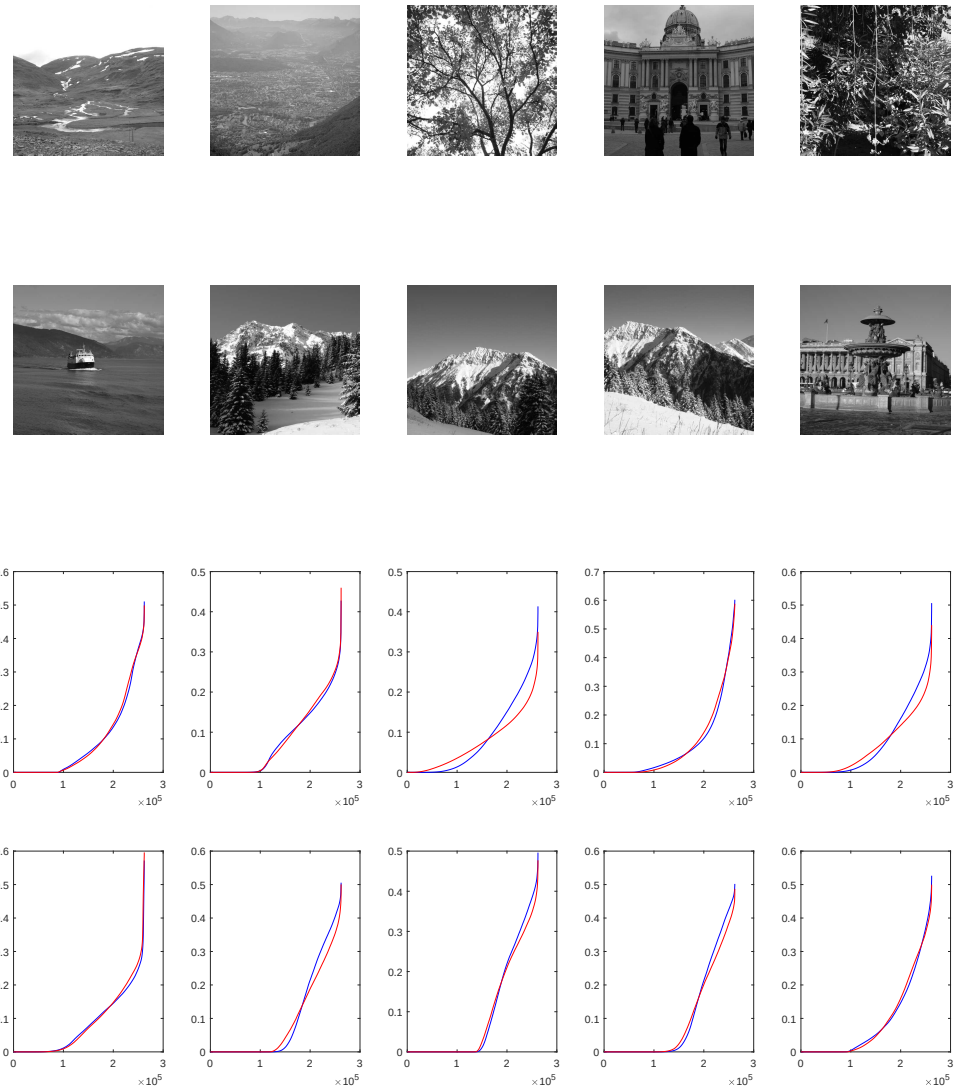


Figure 1. *Sorted total change rate computed for ten images (above) from $\mathcal{B}(1)$ for HILL at 0.4 bpp. Blue: change rates computed from the original images, Red: from noisified covers c_{ij}.*

Experiments

Figure 2 shows the ROCs of three detectors – the optimal LRT (11), the convolutional network SRNet [6], and the maxSRM [12] with the low-complexity linear classifier [9]. The SRNet and the maxSRM were trained on $2 \times 15,000$ images (out of which 2×1000 were used for validation for the SRNet) randomly selected from our dataset and tested on the remaining $2 \times 5,000$ images. The SRNet was first trained on the "easier" dataset $\mathcal{B}(0.5)$ from a random initialization. Then, the trained SRNet was used as a seed for training on $\mathcal{B}(1)$. The LRT's performance was computed on the same testing set. Table 1 shows the comparison in terms of three scalar performance descriptors – the minimal total error probability $P_{\rm E}$ under equal priors

$$P_{\rm E} = \min_{P_{\rm FA}} \frac{1}{2}(P_{\rm FA} + P_{\rm MD}), \tag{15}$$

the false-alarm rate at 50% detection, $P_{\rm FA}(0.5)$, and the missed-detection rate for 5% false alarm, $P_{\rm MD}(0.05)$.

The SRNet is markedly better than the detector with maxSRM. For S-UNIWARD, the gap between the bound and the ROC of the empirical detector was roughly cut by half. For HILL, there seems more room for improvement. In both cases, our analysis shows space for improvement for both embedding algorithms especially for low false alarms.

Justification of design choices

In this section, we provide a justification for the choices made when creating the dataset in Section "Cover source." First, we demonstrate the impact of narrowing the dynamic range of cover images and the effect of flooring the noise variance with σ_0^2 for noisification (the failure to comply with the fine quantization assumption).

Figure 4 left shows the distribution of the normalized LRT $\Lambda^\star$ (11) under $\mathcal{H}_0$ on a dataset created by skipping both Step 3, narrowing the dynamic range, and Step 4, adjusting the variance for noisification, for the alternative hypothesis with stego images embedded with S-UNIWARD at payload 0.4 bpp. The right figure shows the same distribution when including Step 3 but skipping Step 4. The values below the figures are the mean, variance, skewness, and kurtosis. The normalized LRT is much closer to $\mathcal{N}(0,1)$ when the dynamic range is narrowed but its skewness and kurtosis still indicates deviations from a Gaussian distribution.

This deviation is due to the failure of complying with the fine quantization limit needed for the asymptotic result (14) to hold. To further investigate this issue, we selected the BOSSbase image BOSSbase '559.pgm', which contains many pixels with a small noise variance σ_{ij}^2, and executed the following experiment. The image was independently processed $10,000$ times as described in Section "Cover model" with and without flooring the variance with σ_0^2 in Step 4. Figure 5 shows the distribution of $\Lambda^\star$ for the $10,000$ noisifications for both the cover and stego versions of this image with S-UNIWARD at 0.4 bpp. The top two graphs are for the dataset $\mathcal{B}(\infty)$, i.e., when not flooring the variance with σ_0^2 while the bottom two graphs show the effect of enforcing the fine quantization limit with $\sigma_0 = 1$. Notice that when not enforcing the fine quantization limit, under both hypotheses (9) the distribution of the LRT fails to follow (14) – the distribution is asymmetrical with a thicker right tail, which is pronounced especially for the alternative hypothesis. After flooring the variance, however, both distributions become much closer to the expected limit $\mathcal{N}(0,1)$.

Finally, we comment on an alternative to scaling down the dynamic range of the images to resolve the problem with boundary conditions, and that is to skip the scaling and allow the embedding to change pixel values to -1 and 256. Of course, the existence of a single pixel with one of these two values is 100% indicative of embedding. The LRT will "see" this since for such stego pixels the LR will be infinite. However, an empirical detector that forms the test statistic by computing noise residuals or convolutions of the input image may not.

To test this approach, we prepared a different version of our dataset in which the scaling (1) as well as the variance adjustment (2)–(3) were skipped. The detection performance of the LRT for S-UNIWARD at 0.4 bpp is shown in Figure 6 together with the ROC for the SRNet. The third ROC (SRNet-Adjusted) is for the performance of the SRNet with its output augmented by the fact that whenever pixels -1 and 256 are present in the image, it is automatically labeled as stego. The performance of this "informed" SRNet is very close to the theoretical upper bound. Figure 7 shows the distribution of the normalized LRT across cover images. Notice that the distribution indeed matches $\mathcal{N}(0,1)$ even though it is still slightly asymmetric. This is because we do not have the fine quantization assumption satisfied. This measure was not selected for our experiments, however, because of the rather artificial setup as the stego images are not really images.

I am not sure about this experiment as we ignore the fine quantization limit condition here. It would make sense if we floored the variance. Shall we redo?

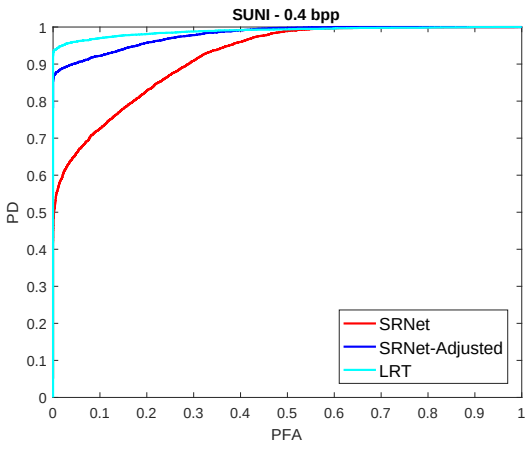


Figure 6. *ROCs when permitting changes to -1 and 256, "SRNet-Adjusted": helping SRNet by pronouncing the test image 'stego' if it contains -1 or 256.*

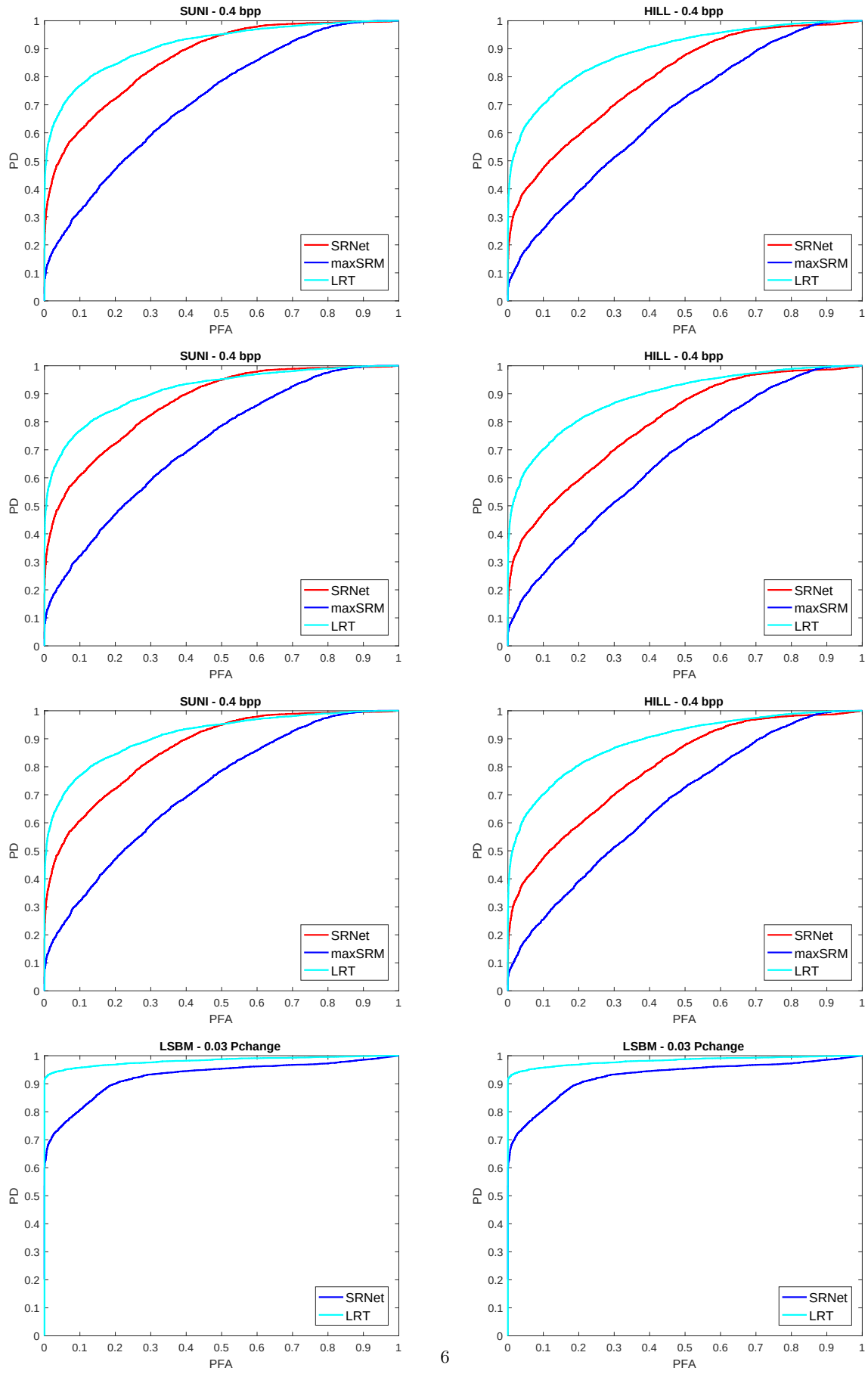


Figure 2. *ROCs for (top down) S-UNIWARD, HILL, WOW at* 0.4 *bpp, and LSBM for total change rate* $\beta = 0.03$ *for the optimal test (LRT), SRNet, and maxSRM with the low-complexity linear classifier with the left and right part columns corresponding to datasets* $\mathcal{B}(0.5)$ *and* $\mathcal{B}(1)$*, respectively.* ***Place holder only.***

	P_{E}				$P_{\mathrm{FA}}(0.5)$				$P_{\mathrm{MD}}(0.05)$			
	HILL	SUNI	WOW	LSBM	HILL	SUNI	WOW	LSBM	HILL	SUNI	WOW	LSBM
SRNet	.2521	.2348										
maxSRM	.2847	.3078										
LRT	.1826	.1611										

Table 1. Performance of three empirical detectors in terms of P_{E}, $P_{\mathrm{FA}}(0.5)$, and $P_{\mathrm{MD}}(0.05)$.

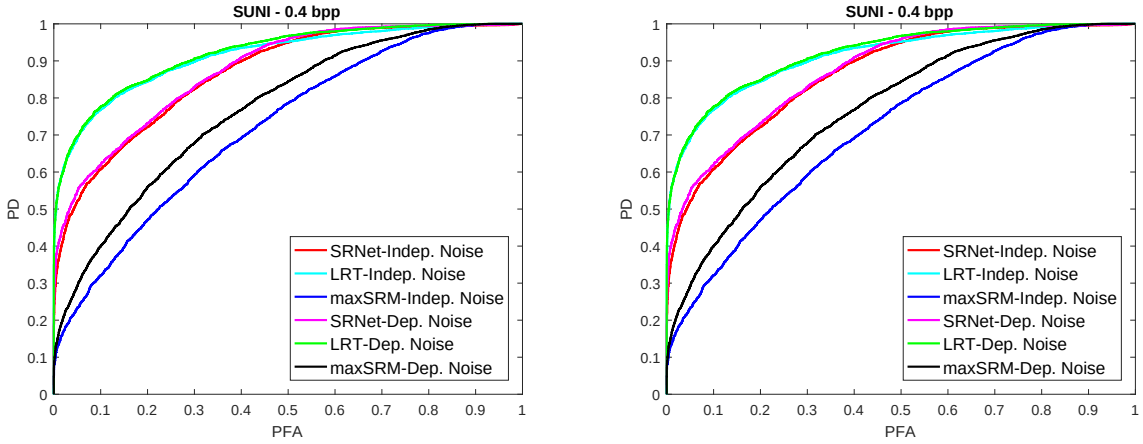
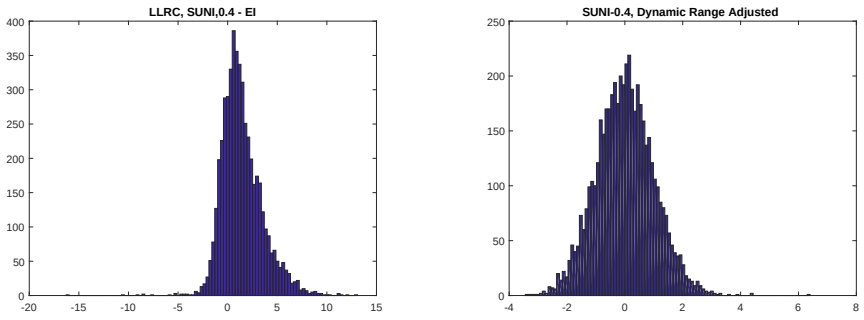


Figure 3. *Contrasting the performance of the LRT and SRNet for payloads 0.2, 0.4, and 0.6 bpp for S-UNIWARD on $\mathcal{B}(0.5)$ and $\mathcal{B}(1)$.* ***Place holder only.***



$(1.44, 4.22, s, k)$ $(0.006, 0.9988, 0.15, 3.44)$

Figure 4. *Left: Distribution of the LRT under $\mathcal{H}_0$ on a dataset created by skipping both Step 3 and 4 for S-UNIWARD 0.4 bpp. Right: The same distribution with Step 3 included but not Step 4. The values below the figures are the mean, variance, skewness, and kurtosis.*

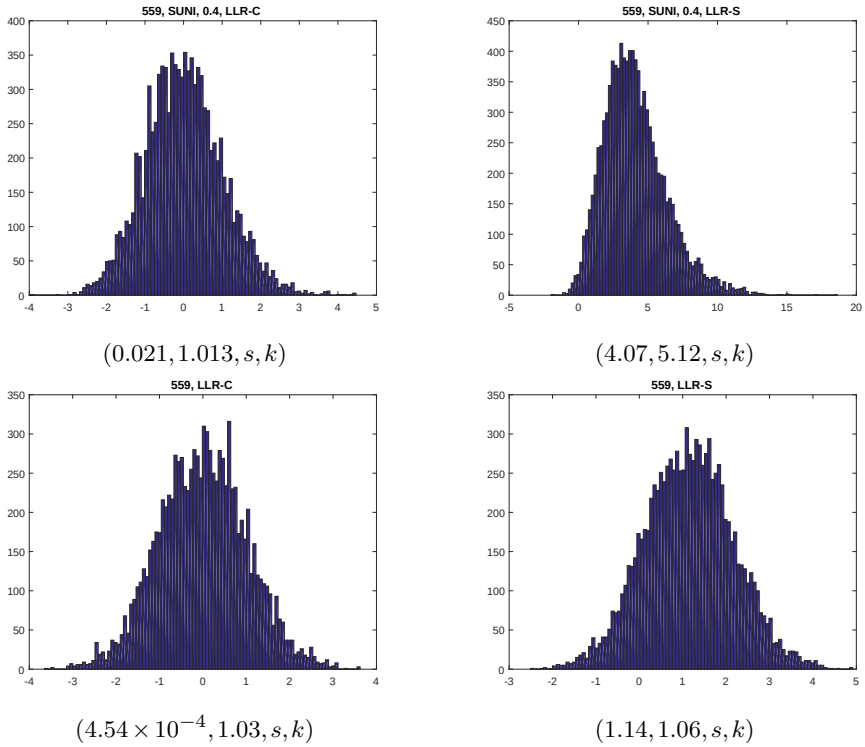


Figure 5. *Distribution of $\Lambda^\star$ and its mean, variance, skewness, and kurtosis for 10,000 different noisifications of BOSSbase image '559.pgm'. Left: covers, Right: stego images for S-UNIWARD 0.4 bpp. Top: dataset $\mathcal{B}(\infty)$, Bottom: $\mathcal{B}(1)$.*

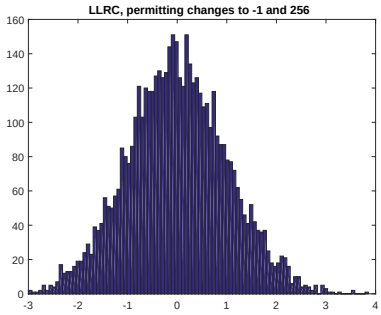


Figure 7. *Distribution of the normalized LRT on cover images across the dataset when permitting changes to -1 and 256. S-UNIWARD at 0.4 bpp.*

Conclusions

TBD

All code used to produce the results in this paper, including the network configuration files are available from `http://dde.binghamton.edu/download/`.

Acknowledgments

The work on this paper was supported by NSF grant No. 1561446 and by DARPA under agreement number FA8750-16-2-0173. 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 DARPA or the U.S. Government.

References

[1] Learning and transferring representations for image steganalysis using convolutional neural network. In *IEEE International Conference on Image Processing (ICIP)*, pages 2752–2756, September 25-28, 2016.

[2] H. Abdulrahman, M. Chaumont, P. Montesinos, and M. Baptiste. Color image steganalysis based on steerable Gaussian filters bank. In F. Perez-Gonzales, F. Cayre, and P. Bas, editors, *The 4th ACM Workshop on Information Hiding and Multimedia Security*, pages 109–114, Vigo, Spain, June 20–22, 2016.

[3] R. J. Anderson and F. A. P. Petitcolas. On the limits of steganography. *IEEE Journal of Selected Areas in Communication*, 16(4):474–481, 1998.

[4] P. Bas, T. Filler, and T. Pevný. Break our steganographic system – the ins and outs of organizing BOSS. In T. Filler, T. Pevný, A. Ker, and S. Craver, editors, *Information Hiding, 13th International Conference*, volume 6958 of Lecture Notes in Computer Science, pages 59–70, Prague, Czech Republic, May 18–20, 2011. Springer Berlin Heidelberg.

[5] P. Bas and T. Furon. BOWS-2. `http://bows2.ec-lille.fr`, July 2007.

[6] M. Boroumand, M. Chen, and J. Fridrich. Deep residual network for steganalysis of digital images. *IEEE*

Transactions on Information Forensics and Security, 2018. To appear.

[7] C. Cachin. An information-theoretic model for steganography. *Information and Computation*, 192(1):41–56, July 2004.

[8] L. Chen, Y.-Q. Shi, and P. Sutthiwan. Variable multi-dimensional co-occurrence for steganalysis. In *Digital Forensics and Watermarking, 13th International Workshop, IWDW*, volume 9023, pages 559–573, Taipei, Taiwan, October 1–4 2014. Springer.

[9] R. Cogranne, V. Sedighi, T. Pevný, and J. Fridrich. Is ensemble classifier needed for steganalysis in high-dimensional feature spaces? In *IEEE International Workshop on Information Forensics and Security*, Rome, Italy, November 16–19 2015.

[10] T. Denemark, M. Boroumand, and J. Fridrich. Steganalysis features for content-adaptive JPEG steganography. *IEEE Transactions on Information Forensics and Security*, 11(8):1736–1746, August 2016.

[11] T. Denemark and J. Fridrich. Improving selection-channel-aware steganalysis features. In A. Alattar and N. D. Memon, editors, *Proceedings IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics 2016*, San Francisco, CA, February 14–18, 2016.

[12] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich. Selection-channel-aware rich model for steganalysis of digital images. In *IEEE International Workshop on Information Forensics and Security*, Atlanta, GA, December 3–5, 2014.

[13] T. Filler, A. D. Ker, and J. Fridrich. The Square Root Law of steganographic capacity for Markov covers. In N. D. Memon, E. J. Delp, P. W. Wong, and J. Dittmann, editors, *Proceedings SPIE, Electronic Imaging, Media Forensics and Security*, volume 7254, pages 08 1–11, San Jose, CA, January 18–21, 2009.

[14] J. Fridrich and J. Kodovský. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, June 2011.

[15] M. Goljan, R. Cogranne, and J. Fridrich. Rich model for steganalysis of color images. In *Sixth IEEE International Workshop on Information Forensics and Security*, Atlanta, GA, December 3–5, 2014.

[16] V. Holub and J. Fridrich. Low-complexity features for JPEG steganalysis using undecimated DCT. *IEEE Transactions on Information Forensics and Security*, 10(2):219–228, February 2015.

[17] A. D. Ker. The square root law of steganography: Bringing theory closer to practice. In R. Bohme and C. Pasquini, editors, *The 5th ACM Workshop on Information Hiding and Multimedia Security*, Philadelphia, June 20–22, 2017.

[18] A. D. Ker. On the relationship between embedding costs and steganographic capacity. In *The 6th ACM Workshop on Information Hiding and Multimedia Security*, Innsbruck, Austria, June 20–22, 2018.

[19] A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich. The Square Root Law of steganographic capacity. In A. D. Ker, J. Dittmann, and J. Fridrich, editors, *Proceedings of the 10th ACM Multimedia & Security Workshop*, pages 107–116, Oxford, UK, September 22–23, 2008.

[20] J. Kodovský and J. Fridrich. Steganalysis of JPEG images using rich models. In A. Alattar, N. D. Memon, and E. J. Delp, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012*, volume 8303, pages 0A 1–13, San Francisco, CA, January 23–26, 2012.

[21] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin. Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters*, 6(12):300–303, December 1999.

[22] P. Moulin and Y. Wang. New results on steganographic capacity. In *Proceedings of the Conference on Information Sciences and Systems, CISS*, Princeton, NJ, March 17–19, 2004.

[23] Y. Qian, J. Dong, W. Wang, and T. Tan. Deep learning for steganalysis via convolutional neural networks. In A. Alattar and N. D. Memon, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2015*, volume 9409, pages 0J 1–10, San Francisco, CA, February 8–12, 2015.

[24] V. Sedighi, R. Cogranne, and J. Fridrich. Content-adaptive steganography by minimizing statistical detectability. *IEEE Transactions on Information Forensics and Security*, 11(2):221–234, 2016.

[25] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang. Steganalysis of adaptive JPEG steganography using 2D Gabor filters. In A. Alattar, J. Fridrich, N. Smith, and P. Comesana Alfaro, editors, *The 3rd ACM Workshop on Information Hiding and Multimedia Security*, IH&MMSec '15, Portland, OR, June 17–19, 2015.

[26] Y. Wang and P. Moulin. Perfectly secure steganography: Capacity, error exponents, and code constructions. *IEEE Transactions on Information Theory, Special Issue on Security*, 55(6):2706–2722, June 2008.

[27] G. Xu. Deep convolutional neural network to detect J-UNIWARD. In R. Bohme and C. Pasquini, editors, *The 5th ACM Workshop on Information Hiding and Multimedia Security*, Philadelphia, June 20–22, 2017.

[28] G. Xu, H.-Z. Wu, and Y. Q. Shi. Ensemble of CNNs for steganalysis: An empirical study. In F. Perez-Gonzales, F. Cayre, and P. Bas, editors, *The 4th ACM Workshop on Information Hiding and Multimedia Security*, IH&MMSec '16, pages 5–10, Vigo, Spain, June 20–22, 2016.

[29] G. Xu, H. Z. Wu, and Y. Q. Shi. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, 23(5):708–712, May 2016.

[30] J. Yang, K. Liu, X. Kang, E. Wong, and Y. Shi. Steganalysis based on awareness of selection-channel and deep learning. In *Internartional Workshop on Digital Forensics and Watermarking*, volume 10431 of *LNCS*, pages 263–272, 2017.

[31] J. Yang, Y.-Q. Shi, E.K. Wong, and X. Kang. JPEG steganalysis based on densenet. *CoRR*, abs/1711.09335, 2017.

[32] J. Ye, J. Ni, and Y. Yi. Deep learning hierarchical representations for image steganalysis. *IEEE*

Transactions on Information Forensics and Security, 12(11):2545–2557, November 2017.

[33] M. Yedroudj, F. Comby, and M. Chaumont. Yedroudj-net: An efficient CNN for spatial steganalysis. Alberta, Canada, April 15–20, 2018.

[34] J. Zeng, S. Tan, B. Li, and J. Huang. Large-scale JPEG image steganalysis using hybrid deep-learning framework. *IEEE Transactions on Information Forensics and Security*, 13(5):1200–1214, 2018.

Author Biography

Mehdi Boroumand received his B.S. degree in electrical engineering from the K. N. Toosi University of Technology, Iran, in 2004 and his M.S. degree in electrical engineering from the Sahand University of Technology, Iran in 2007. From 2007 to 2013 he worked in industry at companies like Ericsson, ZTE, and MTN. He is currently pursuing his Ph.D. degree in Electrical Engineering at Binghamton University. His areas of research include steganography, steganalysis, digital image forensics, and machine learning.

Jessica Fridrich is Distinguished Professor of Electrical and Computer Engineering at Binghamton University. She received her PhD in Systems Science from Binghamton University in 1995 and MS in Applied Mathematics from Czech Technical University in Prague in 1987. Her main interests are in steganography, steganalysis, and digital image forensics. Since 1995, she has received 20 research grants totaling over $12 mil that lead to more than 200 papers and 7 US patents.

Remi Cogranne is an Associate Professor at Troyes University of Technology (UTT), France, since 2013. He has regularly been a visiting scholar at Binghamton University between 2014 and 2017. He received his PhD in Systems Safety and Optimization from UTT in 2011, since on, his research focuses on hypothesis testing applied to image forensics, steganalysis, steganography, and computer network anomaly detection, which lead to more than 60 papers and 3 International patents.