# Detector-Informed Batch Steganography and Pooled Steganalysis

Yassine Yousfi, Eli Dworetzky, and Jessica Fridrich
Binghamton University
Department of Electrical and Computer Engineering
Binghamton, NY 13850
{yyousfi1,edworet1,fridrich}@binghamton.edu

## ABSTRACT

We study the problem of batch steganography when the senders use feedback from a steganography detector. This brings an additional level of complexity to the table due to the highly non-linear and non-Gaussian response of modern steganalysis detectors as well as the necessity to study the impact of the inevitable mismatch between senders' and Warden's detectors. Two payload spreaders are considered based on the oracle generating possible cover images. Three different pooling strategies are devised and studied for a more comprehensive assessment of security. Substantial security gains are observed with respect to previous art – the detector-agnostic image-merging sender. Close attention is paid to the impact of the information available to the Warden on security.

## CCS CONCEPTS

• **Security and privacy**; • **Computing methodologies** → **Image manipulation**; **Neural networks**;

## KEYWORDS

Batch steganography, pooled steganalysis, deep learning, digital image

**ACM Reference Format:**
Yassine Yousfi, Eli Dworetzky, and Jessica Fridrich. 2021. Detector-Informed Batch Steganography and Pooled Steganalysis. In *Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '21), June 22–25, 2021, Virtual Event, Belgium.*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3437880.3460395

## 1 INTRODUCTION

Steganography is the art of hiding information in innocuously looking objects called covers while steganalysis aims to detect evidence that steganography took place. The main bulk of work in this field concerns digital images and focuses on designing embedding algorithms and detectors that perform the best in a single image for a fixed relative payload. In practice, however, the sender can adopt a smarter strategy and distribute the communicated message across multiple covers to decrease the chances of being detected. On the other hand, the Warden is also free to combine evidence from multiple images to decide whether steganography is taking place.

Batch steganography and pooled steganalysis have been originally introduced in [12] together with the so-called shift hypothesis, which claims that the embedding rigidly shifts detector outputs by an amount that depends on the embedded payload size. The first batch strategies [13, 14, 17], which focused on non-adaptive hiding schemes and quantitative detectors, concluded that the payload should either be concentrated in as few covers as possible or spread evenly.

In [14], the author studied pooled steganalysis under the assumption that the Warden knows the chunk sizes but not their assignment to individual images. In a different setup [16], a local outlier factor was used to identify the steganographer from among a large set of users. The topic of learning optimal pooling functions appeared in [**?** ]. Batch steganography with modern content-adaptive embedding algorithms and three ad hoc batch algorithms was studied in [19]. Adversarial embedding [22] was extended to batches of cover images in [18] but performed poorly against an adversarial-aware Warden. In [10], the authors considered batch steganography in JPEG images of different qualities. The optimal size of the bag for Gaussian batch embedding was studied in [21] without considering pooled steganalysis.

The next section explains the reasoning for the setup of batch steganography and pooled steganalysis studied in this paper. To further motivate our work, in Section 3 we demonstrate that the often adopted shift hypothesis is no longer valid for content-adaptive embedding, a fact that holds for the previous generation of detectors built around rich models and linear classifiers as well as modern detectors built as Convolutional Neural Networks (CNNs). In the same section, we show that detectors exhibit highly non-Gaussian distribution on covers. Section 4 contains a formal mathematical description of three pooled detectors considered in this paper. Two novel detector-informed batch steganographic techniques are described and theoretically analyzed in Section 5. The setup of our experiments, including implementation details, is explained in Section 6. The results of all experiments together with their interpretation and discussion appear in Section 7. The paper is concluded in Section 8.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
*IH&MMSec '21, June 22–25, 2021, Virtual Event, Belgium.*
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8295-3/21/06…$15.00
https://doi.org/10.1145/3437880.3460395

## 2 BASIC SETUP

In batch steganography, two actors, Alice and Bob, exchange messages hidden in digital images. To avoid being detected by an adversary (the Warden), they use modern content-adaptive spatial-domain steganography and adjust the payload size embedded in each image to decrease the risk of being detected. The Warden combines the outputs of a single-image detector applied to all images exchanged by Alice and Bob in an effort to discover the use of a steganographic channel and not necessarily identify which images are cover and stego.

This problem of batch steganography and pooled steganalysis may accept many different forms depending on what information about the cover source, the steganographic method, the payload spreading strategy, and possibly Warden's detector is available to all actors. Following Kerckhoffs's principle, we are mainly interested in the situation when the Warden has full knowledge of algorithms used by Alice and Bob but not any shared secret or specific data used by the senders. In particular, we assume that the steganographers and the Warden have access to the same source of covers, which they can use in any way to design a payload spreading strategy as well as build detectors. We will also assume that the Warden knows the steganographic method that might be in use and the payload-spreading strategy. For example, if the steganographers use feedback from a detector to determine the size of payload chunks embedded in each image, the Warden can train the same detector architecture on her end but it will ultimately be a slightly different detector because of different training data. Moreover, the payload chunk sizes will also generally depend on the cover images to which the Warden does not have access.

Having said this, we will at times consider a payload-aware Warden that has access to the exact payload chunk sizes that Alice sends as a form of a worst case scenario and to evaluate the impact of the lack of such precise knowledge.

While the steganographers may opt for a spreading strategy that is free of any assumptions about Warden's detector, such as the Image Merging Sender (IMS) and Detectability / Distortion Limited Senders (DeLS / DiLS) considered in [19], they are free to guess and make use of a detector that is likely to be used by the Warden or any other detector. The specific assumptions made in this paper will be clarified later based on discussions and other important experimental facts concerning content-adaptive embedding and modern steganalysis detectors.

## 3 NEW CONTEXT

The problem of batch steganography and pooled steganalysis has been revisited many times throughout the history of this field. In this section, we challenge some of the assumptions made in prior art to motivate our approach.

In [19], an argument based on the Central Limit Theorem (CLT) was made that, on cover images, the outputs of a single-image detector that uses high-dimensional rich models and a linear classifier is zero-mean Gaussian. Leveraging the shift hypothesis, the authors further assumed that the embedding merely shifts this distribution by an amount that depends on the embedded payload. The Gaussianity and the shift hypothesis allowed the authors to derive an optimal pooled detector in the form of a matched filter, which in practice correlates detector outputs with shifts estimated from near embedding invariants and the payload itself. Within this context, they studied the IMS and DeLS (DiLS), the last two spreading so that the same level of detectability (distortion) is induced in every image.

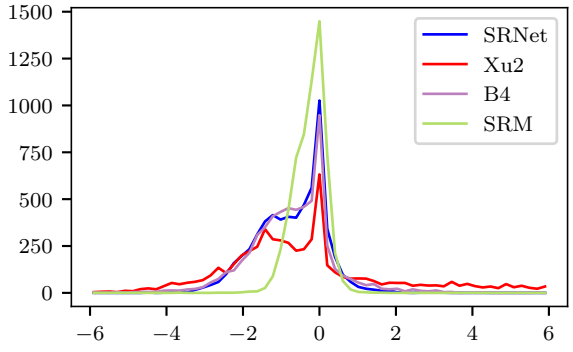


**Figure 1: Distribution of the soft detector output for SRNet, EfN B4, Xu2, and SRM with LCLC trained on covers vs. uniform payload mixture of HILL.**

Below, we demonstrate that modern detectors not only exhibit highly non-Gaussian behavior but also clearly fail to satisfy the shift hypothesis. This is true for both non-adaptive and content-adaptive steganography and detectors based on rich models as well as CNNs. For better readability, the description of datasets and detectors, including their training for all experiments commented upon in this section is postponed to Section 6.

### 3.1 Non-Gaussian distribution on covers

Figure 1 shows the distribution of soft outputs of four detectors on $256 \times 256$ grayscale cover images from ALASKA II when training them as binary detectors on cover versus stego images embedded with a uniform mixture of payloads from {0.05, 0.1, 0.2, . . ., 1.4, 1.5} bpp. The soft output for the Spatial Rich Model (SRM) [8] implemented with the Low Complexity Linear Classifier (LCLC) [4] is the projection of the rich feature on the weight vector. For the three CNNs, SRNet [2], Efficient Net B4, and SE-ResNet18 (Xu2 net), the output is the logit. The cover distribution for all detectors is highly asymmetric and spiky. The distribution on covers is also clearly non-Gaussian and bimodal for the networks with the left "hump" corresponding to "easy covers."

While the fact that CNNs produce highly non-Gaussian outputs on both cover and stego images is less surprising due to their inherent non-linearity, rich model features are also non-linear functions of the image since they are higher-order statistics (histograms) of quantized and truncated noise residuals.

### 3.2 Failure of the shift hypothesis

Figure 2 shows the distribution of two of the above four detectors on stego images embedded with a range of fixed relative payloads. With increased payload size, the distribution gradually "morphs" to the right, affecting mostly the distribution tails, while the peak at zero stays nearly stationary. In fact, in order to obtain a rigidly

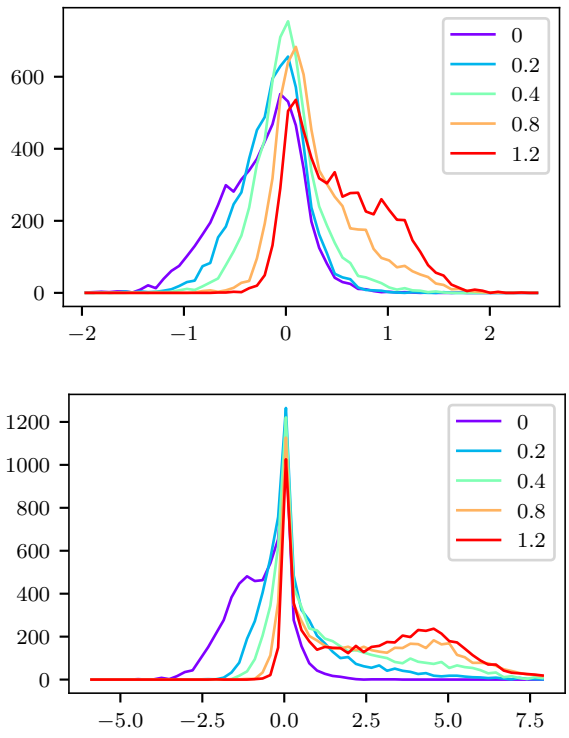


**Figure 2: Distribution of detectors' soft output on cover and stego images embedded with HILL for a fixed relative payload. Top: LCLC with SRM, bottom: SRNet. Note that the distributions morph in a far more complex manner than a simple shift.**

shifted distribution, one would need to adopt a non-trivial spreading strategy (see Section 5). The shift hypothesis, as originally conceived in [12], is likely limited to quantitative detectors since their expected test statistic is the change rate (payload).

### 3.3 Complex response curves

Undoubtedly, the key element in considering the problem of spreading payload across images is the response of the Warden's detector as a function of the payload size – the detector's response curve – which depends on the cover image and the steganographic method. A cover image with a completely flat response curve would be ideal for embedding a large payload as the embedding is "invisible" to the detector. And this is true regardless of whether it is detected as cover or stego. On the other hand, an image exhibiting a steep response curve should hold a comparatively smaller payload.

Since embedding a secret message is a stochastic process, the detector response naturally exhibits a statistical spread, which increases with increased payload (see Figure 3). To eliminate this source of randomness, we define the response curve (RC) for a given cover image and detector as the expected value of the response over embeddings with different stego keys (seeds for an embedding simulator). In Figure 3, the RCs are rendered with thick blue lines obtained by averaging over 100 embeddings for each payload with the light blue shade used to depict the standard deviation.

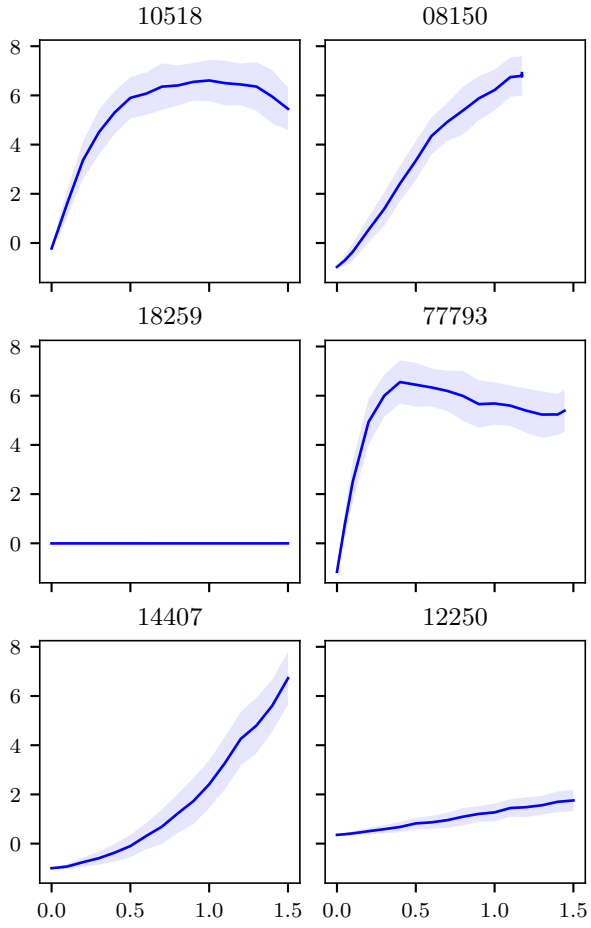


**Figure 3: SRNet's response curves (the logit as a function of embedded payload size) for six selected images. The solid line is the expectation over 100 embeddings with different stego keys with the light blue shade used to depict the standard deviation. Detector: SRNet trained on uniform payload mixture for HILL.**

The diversity of these responses is responsible for the failure of the shift hypothesis.

Note that RCs are mostly non-decreasing with the exception of a few images for which the response decreases for very large payloads (e.g., image 10518). Despite the slight drop, the final class label is unlikely to flip because the logit values are still very large. While we are not certain why this is happening, it might be due to the fact that the content-dependent stego noise for large payloads might start resembling sensor noise in some images. To simplify our reasoning, we adopt the feasible assumption that all RCs are non-decreasing for the entire payload range.

The RCs tell the tale of what is happening at detection. For image 14407, the RC is initially flat and then steeply bends upwards. This is likely because the image contains some complex content where the detection is difficult, and once the embedding "spills over" to other parts of the image due to increased payload, it quickly starts contributing to detectability. The flat curves of images 18259 and

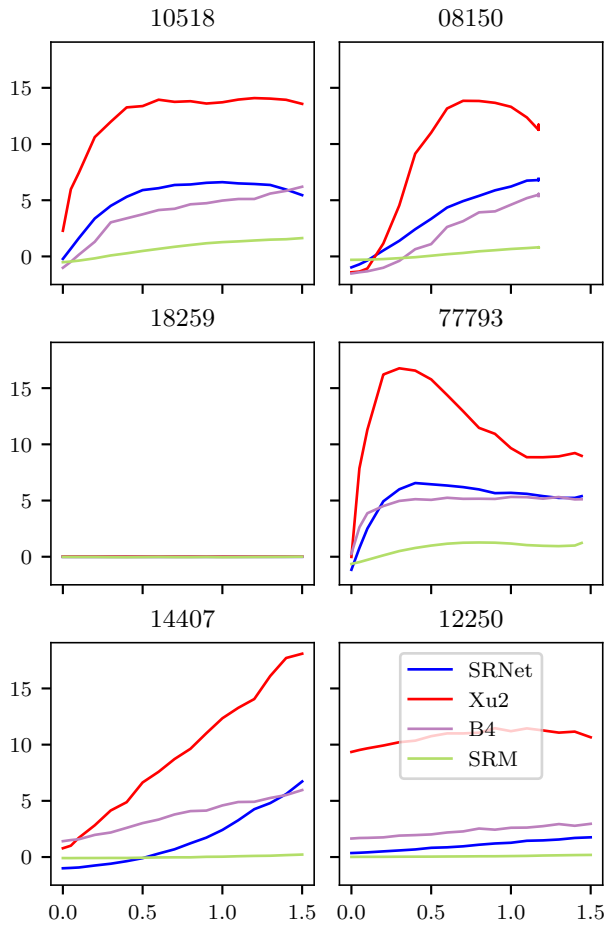


**Figure 4: Response curves for the same six selected images as in Figure 3 and four different detectors.**

12250 mean that they can hold a very large payload without changing the detector's output. Lastly, we point out the steep response curves for images 10518, 77793, and 08150 with smooth content where embedding quickly becomes very detectable. Note that for image 08150, the maximal embeddable payload is only about 1.2 bpp because the image contains wet pixels [7].

Figure 4 shows the RCs for the same six images for four different detectors. Note that while the network detectors are very different deep architectures, the response curves exhibit qualitative similarities. This justifies our choice to use detector output as feedback for batch steganography.

Finally, we remark that the steganographers must select their images randomly from their cover source as any cover selection or rejection would skew the statistics of the cover source, a fact that would be detected by the Warden who is testing whether her detector's outputs are consistent with the detector distribution on covers. Thus, the best the sender can do is to minimize the disturbance to the distribution of Warden's test statistic. We come back to this problem in the next section when we lay out a more detailed formulation of our setup.

## 4 POOLED STEGANALYSIS

In this section, we describe three pooling strategies for the Warden that will be used to assess security of batch steganography in this paper.

We will assume that the steganographers maintain an average payload per pixel $r \in [0, \log_2 3]$, the communication rate. By the square root law [15], this means that, asymptotically, they will be caught with near certainty. Our goal is not perfect or bounded security, which would require the communication rate to taper off to zero, but to minimize the detectability in each bag of images. For simplicity, in the rest of this paper we assume that the Warden knows $r$ and that the embedding method is fixed and known to all actors.

Let $\mathcal{X}$ denote the set of all possible cover images of some fixed size. A cover bag of size $B$, $\mathbf{X} = (X_0^{(1)}, \dots, X_0^{(B)})$, is formed by selecting $B$ cover images $X_0^{(1)}, \dots, X_0^{(B)} \in \mathcal{X}$ according to some probability distribution over $\mathcal{X}$. A spreading strategy $S$ induces a unique mapping $\alpha_{r,S} : \mathcal{X}^B \to [0, \log_2 3]^B$ that determines the relative payloads (in bpp) embedded in the $B$ images using a ternary steganographic scheme. When $r$ and $S$ are clear from context, we simply write $\alpha_i \in [0, \log_2 3]$ to denote the $i$th component of $\alpha_{r,S}(\mathbf{X})$, i.e., the relative payload embedded in the $i$th image. The map $\alpha_{r,S}$ must satisfy the payload constraint

$$\sum_{i=1}^{B} \alpha_i = rB. \tag{1}$$

A payload tag for rate $r$ is the relative payload $\tau_{r,S}(X_0^{(i)})$ that the $i$th image receives for an infinitely large bag.

A single-image detector is a mapping $d : \mathcal{X} \to \mathbb{R}$ that assigns to each image a soft output from the detector. The soft outputs can be thresholded for a hard cover / stego decision based on application-dependent requirements, such as controlling the false alarm rate. The response curve for image $X_0^{(i)}$ and detector $d$ is the function $\varrho_i : [0, \log_2 3] \to \mathbb{R}$

$$\varrho_i(\alpha) = \mathbb{E}[d(X_\alpha^{(i)})] \tag{2}$$

obtained as the expected value of $d$ on stego images $X_\alpha^{(i)}$ when embedding cover $X_0^{(i)}$ with payload $\alpha$ with random messages and stego keys. To distinguish the mathematical objects used by the Warden from those used by the steganographers, we will use the superscript 'W' for the Warden and 'S' for the steganographers. Pooled detectors will be denoted with the Greek letter $\pi$.

Let $f_0^{\mathrm{W}}$ denote the Warden's detector distribution on covers (c.f., Figure 1). Given $B$ images $Y^{(i)}$, $i = 1, \dots, B$, communicated by the sender and under inspection by the Warden, $\mathbf{Y} = (Y^{(1)}, \dots, Y^{(B)})$, the Warden computes $d^{\mathrm{W}}(Y^{(i)})$ for all images and infers whether the sender uses steganography. In the absence of any other knowledge about the spreading strategy or the communication rate, the Warden would face a composite hypothesis test, namely testing for a known distribution:

$$\begin{aligned} \mathcal{H}_0 &: \quad d^{\mathrm{W}}(Y^{(i)}) \sim f_0^{\mathrm{W}} \quad \text{for all } i \\ \mathcal{H}_1 &: \quad d^{\mathrm{W}}(Y^{(i)}) \nsim f_0^{\mathrm{W}} \quad \text{for some } i. \end{aligned} \tag{3}$$

### 4.1 Correlator pooling

Since we assume that the Warden knows the spreading strategy and the communication rate $r$, she can test for an increase in the detector response $s_i = \varrho_i^{\mathrm{W}}(\alpha_i) - \varrho_i^{\mathrm{W}}(0)$. However, since she does not have access to cover images, she needs to estimate the response on the cover, $\varrho_i^{\mathrm{W}}(0)$, or simply add it to the modeling error. Moreover, a realistic Warden will also need to estimate the payloads $\alpha_i$ from the images at hand. In particular, she can obtain the estimated payload $\hat{\alpha}_i$ possibly embedded in $Y^{(i)}$ by computing the $i$th component of $\alpha_{r,S}(\mathbf{Y})$. The statistical hypothesis testing problem thus becomes

$$\begin{aligned} \mathcal{H}_0: &\quad d^{\mathrm{W}}(Y^{(i)}) = \xi_i && \text{for all } i \\ \mathcal{H}_1: &\quad d^{\mathrm{W}}(Y^{(i)}) = \hat{s}_i + \xi_i && \text{for all } i, \end{aligned} \tag{4}$$

where $\hat{s}_i = \hat{\varrho}_i^{\mathrm{W}}(\hat{\alpha}_i) - \hat{\varrho}_i^{\mathrm{W}}(0)$ is the estimated expected increase of the detector output using a RC $\hat{\varrho}_i^{\mathrm{W}}(\alpha)$ computed from the image at hand $Y^{(i)}$, $\hat{\varrho}_i^{\mathrm{W}}(\alpha) = \mathbb{E}[d^{\mathrm{W}}(Y_\alpha^{(i)})]$, where $Y_\alpha^{(i)}$ is image $Y^{(i)}$ embedded with relative message $\alpha$, and $\xi_i$ is the modeling error.

In the simplest case of independent Gaussian noise samples $\xi_i \sim \mathcal{N}(0, \sigma_i^2)$, the optimal detector would be the generalized matched filter (correlator) [11]. In our work, we experimented with several different forms of the estimators, including the estimator used in [19] that used near embedding invariants. In the end, the best overall performance was achieved with a pooled detector in the form of a correlator

$$\pi_{\mathrm{COR}}(\mathbf{Y}) = \sum_{i=1}^{B} d^{\mathrm{W}}(Y^{(i)})\gamma(\hat{\alpha}_i), \tag{5}$$

where $d^{\mathrm{W}}(Y^{(i)})$ are detector outputs on the analyzed images and $\gamma(\alpha)$ is a logistic fit to embedding shifts $\hat{\varrho}_i^{\mathrm{W}}(\alpha) - \hat{\varrho}_i^{\mathrm{W}}(0)$ across a dataset of cover images $i$.

### 4.2 LRT pooling

Another possibility for the Warden is to test whether the detector output for the $i$th image is consistent with the distribution of her detector $f_{\hat{\alpha}_i}^{\mathrm{W}}$ on stego images all embedded with the same relative payload $\hat{\alpha}_i$:

$$\begin{aligned} \mathcal{H}_0: &\quad d^{\mathrm{W}}(Y^{(i)}) \sim f_0^{\mathrm{W}} && \text{for all } i \\ \mathcal{H}_1: &\quad d^{\mathrm{W}}(Y^{(i)}) \sim f_{\hat{\alpha}_i}^{\mathrm{W}} && \text{for all } i \end{aligned} \tag{6}$$

with the optimal pooled detector being the log-likelihood ratio

$$\pi_{\mathrm{LRT}}(\mathbf{Y}) = \sum_{i=1}^{B} \log \frac{f_{\hat{\alpha}_i}^{\mathrm{W}}\left(d^{(w)}(Y^{(i)})\right)}{f_0^{\mathrm{W}}\left(d^{(w)}(Y^{(i)})\right)}. \tag{7}$$

### 4.3 Tag based pooling

We also consider the pooling strategy where the Warden makes use of the tags $\tau_i = \tau_{r,S}(X_0^{(i)})$ and trains her single-image detector as a binary classifier between covers and stego images embedded with tags. For large enough bags, this is the correct stego source from which the sender draws their images. Note that the stego source only depends on the spreading strategy $S$ and rate $r$. We make an argument that, for large bags and for images in the bags selected randomly, the optimal pooling strategy is averaging the detector logits. This is because all three deep learning architectures used in this paper apply average pooling[1] in the last convolutional layer before the fully connected layer. If averaging the detector outputs across all images in the bag was not the best strategy, one could obtain a better single-image detector by splitting each image into smaller tiles, applying the network to the tiles and learning a more sophisticated strategy for combining the outputs. In summary, for a tag-based single-image detector $t^{\mathrm{W}}$, our pooling strategy is

$$\pi_{\mathrm{TAG}}(\mathbf{Y}) = \frac{1}{B}\sum_{i=1}^{B} t^{\mathrm{W}}(Y^{(i)}). \tag{8}$$

### 4.4 Average pooling

As the last option considered in this paper, we added a fourth baseline pooling strategy in the form of a simple average of detector outputs on analyzed images $Y^{(i)}$:

$$\pi_{\mathrm{AVG}}(\mathbf{Y}) = \frac{1}{B}\sum_{i=1}^{B} d^{\mathrm{W}}(Y^{(i)}). \tag{9}$$

We also experimented with the max pooling strategy $\pi_{\mathrm{MAX}}(\mathbf{Y}) = \max_i d^{\mathrm{W}}(Y^{(i)})$ but do not report on it because it performed very poorly w.r.t. the other strategies.

Note that in this paper, we will consider both a payload-aware Warden that knows the senders' payloads $\alpha_i$ as well as a realistic Warden that needs to estimate both from the image at hand.[2]

## 5 BATCH STEGANOGRAPHY

In this section, we describe two types of detector-informed spreading strategies depending on the adopted statistical model for the cover source. We also provide theoretical analysis of both senders under certain simplifying assumptions. This analysis will be used to better understand and interpret the experimental results in Section 7. The sender's single-image detector will be denoted $d^{\mathrm{S}}$.

### 5.1 Shift limited sender

The Shift Limited Sender (SLS) enforces the shift hypothesis by considering the impact of the embedding on the statistical distribution of detector outputs across cover images. To embed an average communication rate $r$ in $B$ cover images $X_0^{(i)}$, the SLS sender finds the smallest $\delta > 0$ that leads to the same expected detector output shift when embedding payload $\alpha_i$ in $X_0^{(i)}$, satisfying $\sum_{i=1}^{B}\alpha_i = rB$, and

$$\delta = \varrho_i^{\mathrm{S}}(\alpha_i) - \varrho_i^{\mathrm{S}}(0) \tag{10}$$

for all $i$ for which $\varrho_i^{\mathrm{S}}\left(\alpha_{\max}(X_0^{(i)})\right) - \varrho_i^{\mathrm{S}}(0) \geq \delta$, where $\alpha_{\max}(X_0^{(i)}) \leq \log_2 3$ is the maximal embeddable payload in $X_0^{(i)}$ equal to the relative number of non-wet pixels. For images that do not satisfy this condition (images with flat response curves), we set $\alpha_i = \alpha_{\max}(X_0^{(i)})$.

As explained in Section 6 in more detail, the SLS was implemented numerically using unidirectional search for $\delta$ with the image response curves.

[1] The word 'pooling' not to be confused with pooling as used in pooled steganalysis.
[2] More on this appears in Section 7.2.

To obtain better insight, below we derive a closed form for the payload by adopting a linear model for response curves:

$$\varrho_i^{\mathrm{S}}(\alpha_i) - \varrho_i^{\mathrm{S}}(0) = b_i\alpha_i, \tag{11}$$

with $b_i > 0$. This means that we essentially assume that the RCs are not completely flat, and we ignore the upper bound on $\alpha_i \leq \alpha_{\max}(X_0^{(i)})$.

Since the SLS requires $b_i\alpha_i = \delta$ for all images $X^{(i)}$ in the bag, the payload constraint (1) implies that $\delta = rB/\sum_{i=1}^{B} 1/b_i$, which gives us the following expression for $\alpha_i$

$$\alpha_i = \frac{rB}{b_i \sum_{k=1} \frac{1}{b_k}}. \tag{12}$$

## 5.2 Minimum deflection sender

The Minimum Deflection Sender (MDS) considers a statistical model for *each scene* rather than across images. The specific cover used by the sender is a sample from an acquisition oracle taking images of the same scene with the same acquisition device. Sensor noise and possibly small spatial shifts and rotations due to camera shake would contribute to the randomness.

The main advantage of this approach is that the detector output on such cover images is well modeled by a Gaussian distribution due to the fact that the detector can be linearized on the neighborhood of the noise-free cover image. We assume that the embedding changes the expectation of the detector output based on the response curve but does not change the variance. Hence, the sender determines the payloads to minimize the power of the most powerful detector for the following hypothesis testing problem:

$$\begin{aligned} \mathcal{H}_0 &: \quad d^{\mathrm{S}}(Y^{(i)}) \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad \text{for all } i \\ \mathcal{H}_1 &: \quad d^{\mathrm{S}}(Y^{(i)}) \sim \mathcal{N}(\mu_i + s_i, \sigma_i^2) \quad \text{for all } i, \end{aligned} \tag{13}$$

where $\mu_i$ is the expected value of $d^{\mathrm{S}}$ on cover images generated by the acquisition oracle for the $i$th image and $s_i$ is the expected increase of detector response due to embedding payload $\alpha_i$. Note that in (13) we assume that the acquisition variance dominates the variance due to embedding a random message, hence the variances are equal under both hypotheses. For a clairvoyant Warden who uses the same detector $d^{\mathrm{W}} = d^{\mathrm{S}}$ and knows $\mu_i$ and $\sigma_i^2$, with cover images drawn independently from the cover source, the most powerful detector is the likelihood ratio test, which assumes the form of a mean-shifted Gauss-Gauss problem. Thus, its performance is determined by the deflection coefficient $\sum_{i=1}^{B} s_i^2/\sigma_i^2$.

For practical implementation, we will assume that $d^{\mathrm{S}}(X_0^{(i)}) = \varrho_i^{\mathrm{S}}(0) \approx \mu_i$ is approximately equal to the expected detector output across all acquisitions. Hence, the MDS selects the $\alpha_i$ to be embedded in $X_0^{(i)}$ that minimizes the deflection[3]

$$\Delta^2(\mathbf{X}) = \sum_{i=1}^{B} \frac{\left(\varrho_i^{\mathrm{S}}(\alpha_i) - \varrho_i^{\mathrm{S}}(0)\right)^2}{\sigma_i^2}. \tag{14}$$

While our assumptions about Warden's access to $d^{\mathrm{S}}$ and $\mu_i$ and $\sigma_i^2$ are too idealistic, we can claim that the MDS considers the worst case scenario. Since estimating the variances $\sigma_i^2$ experimentally using the acquisition oracle would be far too elaborate and even infeasible in many cases, we further simplify the MDS by assuming that the variances $\sigma_i^2$ are all approximately the same. Experiments on Monobase [1] with simulated acquisition at higher ISO (as in Natural Steganography [1]) confirmed that the detector variance is indeed rather stable across different scenes.

To obtain insight into how the MDS assigns payloads, we again derive a closed form expression for $\alpha_i$ using the linear model (11) for the response curves $\varrho_i^{\mathrm{S}}(\alpha_i) - \varrho_i^{\mathrm{S}}(0) = b_i\alpha_i$. To minimize the deflection $\Delta^2(\mathbf{X})$ with equal variances $\sigma_i^2 = \sigma^2$ subject to the payload constraint (1), we find the stationary point of the Lagrangian

$$\mathscr{L} = \frac{1}{2}\sum_{k=1}^{n} b_k^2\alpha_k^2 - \lambda\left(\sum_{k=1}^{n}\alpha_k - rB\right), \tag{15}$$

which yields the closed form for MDS payloads

$$\alpha_i = \frac{rB}{b_i^2 \sum_{k=1}^{B} \frac{1}{b_k^2}}. \tag{16}$$

# 6 IMPLEMENTATION

In this section, we list the details regarding our implementation of the batch steganography algorithms as well as the detectors.

## 6.1 Datasets

The dataset is the ALASKA II split into three parts (Split 1, 2, and 3), each containing 25,000 images further split into 22k, 1k, and 2k images for training, validation, and testing. The splits are used to study the impact of a mismatched training set for training Warden's detector. The images were developed as in [5] without the final JPEG compression step. Alice uses the test set of Split 1 to send her secret messages in bags of size $B$ by sampling $B$ images with replacement.

Because of the sheer amount of possible combinations of the steganographer's detector, the Warden's detector, stego schemes, communication rates $r$, bag sizes, and spreading / pooling strategies, we limit our exposition to the steganographic scheme HILL[4] and mainly the rate $r = 0.3$ bpp. Instead of reporting the complete set of results for all possible setups, we highlight the most interesting and relevant findings.

## 6.2 Single-image detectors

For spreading, the sender uses a single-image detector $d^{\mathrm{S}}$ in the form of an SRNet (SRNet1) trained on Split 1. Splits 2 and 3 are used by the Warden who will train $d^{\mathrm{W}}$ as another instance of SRNet (SRNet2) on Split 2, Xu2 on Split 2, EfN B4 on Split 3, and SRM on Split 3. EfN B4 and Xu2 were modified by removing the average pooling and strides from the first two layers as described in [23]. All network detectors are pre-trained on ImageNet, SRNet was pre-trained on a binary task of steganalyzing J-UNIWARD [9] (the so-called JIN pre-training exactly as described in [3]), while the other networks were pre-trained on the ImageNet classification task.[5] Steganalysis training on HILL / MiPOD is done with relative

[3] As explained in Section 6, for the MDS we use a logistic fit to the RCs instead of the RCs to allow for a more efficient gradient descent based search algorithm.

[4] In particular, since we observed qualitatively and quantitatively similar conclusions for MiPOD, the results are not reported.

[5] Downloaded from https://github.com/rwightman/pytorch-image-models

payloads randomly drawn from the uniform distribution on the set of relative payloads $\mathcal{P} = \{0.05, 0.1, 0.2, \ldots, 1.4, 1.5\}$.

We also add another, qualitatively different single-image detector based on the Spatial Rich Model (SRM) [8] and the LCLC, also trained on payloads randomly uniformly drawn from $\mathcal{P}$.

### 6.3 Pooled detectors

For the correlator pooling strategy, the Warden uses her test set to fit a logistic curve to the embedding shifts $\varrho_i^{\mathrm{W}}(\alpha) - \varrho_i^{\mathrm{W}}(0)$ to obtain $\gamma(\alpha)$. The logistic curve is defined as

$$p(x) = \frac{a}{1 + e^{c(x-m)}} + h, \tag{17}$$

with $0 < a, m < \infty$, $-\infty < c < 0$, $h \in \mathbb{R}$, and the fit is performed using non-linear least squares[6] initialized at $(a, m, c, h) = (1, 1, -1, 0)$.

For the LRT pooling strategy, the Warden embeds her test set with a set of relative payloads $\mathcal{P} = \{0.05, 0.1, 0.2, \ldots, 1.4, 1.5\}$. Then she proceeds to estimate the distribution of the detector's output $f_\alpha^{\mathrm{W}}$ for each $\alpha \in \mathcal{P}$.[7] To cover the entire range of possible payloads, the Warden linearly interpolates between likelihoods evaluated at the payload grid $\mathcal{P}$.

For the tag-based poolers, the Warden fine-tunes her single image detectors on a dataset embedded with tags computed by randomly grouping all training images into bags of $B = 100$. Note that the Warden has to train a tag-based pooler for each spreading strategy and average communication rate.

### 6.4 Senders

The IMS was implemented by considering a given bag of $B$ images each with $N$ pixels as a single large image into which the total payload of $rBN$ bits was embedded using an embedding simulator. The costs were pre-computed from single images. We would like to point out that this version of the IMS differs from the implementation used in [19]. There, the authors first pre-computed tags for all images from their dataset and then simply selected $B$ images for a given bag. Thus, the communication rate $r$ varied from bag to bag, and was maintained across bags only in expectation. This difference is rather important as will become apparent when studying the detectability as a function of $B$.

The SLS was implemented by searching for the smallest $\delta$ satisfying (10) using unidirectional search. The SLS uses the RCs estimated from 100 embeddings of the cover image as explained in Section 3.3, and linearly interpolates between grid points to cover the entire range of possible payloads.

The MDS makes use of the same logistic model as in (17), fit to each RC. A projected gradient descent with momentum initialized with IMS payloads for each bag was used to search for the payloads that minimize the deflection (14). To facilitate convergence, the learning rate and momentum were updated according to a one-cycle scheduler [**?** ]; the learning rate and momentum fluctuated within the intervals $[10^{-2}, 10^{2}]$ and $[.90, .99]$, respectively. To comply with the payload constraint and bounds $0 \le \alpha_i \le \alpha_{\max}(X^{(i)})$, at each step of the gradient descent the vector of payloads was projected to the feasible set of points, a hyperplane formed by (1) contained within the $B$-dimensional box $[0, \alpha_{\max}(X^{(1)})] \times \ldots \times [0, \alpha_{\max}(X^{(B)})]$.

[6] Using scipy's curve_fit function
[7] Using scipy's gaussian_kde function

## 7 EXPERIMENTS

This section contains the results of all our experiments and their discussion. In particular, the proposed detector-informed senders are evaluated against the IMS with four pooling strategies. Substantial space is devoted to studying the impact of the information available to the Warden as well as the effect of Warden's choices for the single-image detector.

### 7.1 Best spreading and pooling strategies

In this section, we compare the SLS and MDS and the previously proposed IMS. We also evaluate all poolers to see which pooling strategy is the best. We do so for a range of bags and one fixed setup with $r = 0.3$ bpp and HILL. The Warden uses the same architecture as the senders, the SRNet, trained on Split 2 (SRNet2) because it is not feasible to assume that the Warden has the same training set. In this section, we give the Warden the exact payloads $\alpha_{r,S}(\mathbf{X})$ that might be embedded in each bag. In reality, the Warden would have to estimate the payloads for each bag, which is likely to decrease the detectability. We simplify here because executing experiments at scale with having to estimate the payloads is very time consuming as the Warden needs to estimate the average response curves w.r.t. her detector for all images in the bag. In Section 7.2, we show that the effect of using the estimated payloads leads to only a small drop in detection accuracy and thus does not affect the results or our conclusions much.

The detection performance of pooled detectors is reported using the weighted Area Under the ROC Curve (wAUC) as used in ALASKA II [5]. We note that the pooled detector makes a binary decision about each bag being either cover or stego. Figure 5 shows the wAUC of four different poolers versus the bag size. Both detector-aware senders offer much better security when compared to the detector-agnostic IMS.

Note that for all senders, as the bag size grows, the detectability initially decreases and eventually starts increasing due to the Square Root Law since the senders maintain a positive communication rate $r$. The initial drop, which is far more pronounced for the two detector-aware senders, can be explained by considering the response curves. If a bag contains an image with a nearly flat response curve, it will be embedded close to its maximum capacity while other images will receive smaller payloads. Taking a bag of two as an example, it is more advantageous for the sender to embed payload 0.6 bpp in one of the images rather than 0.3 in each. The spreading thus initially helps decrease detectability to a point when the SRL starts engaging and the bags provide more data to reach a more reliable decision about the use of steganography. Note that this result is in stark contrast with the behavior of the IMS from [19] because the IMS there worked with fixed tags attached to all images and only embedded a given relative payload in each bag *on average*. Thus, it was unable to utilize the effect discussed above. Our concept of batch steganography in bags is more flexible and makes better use of the available cover images especially for small bags.

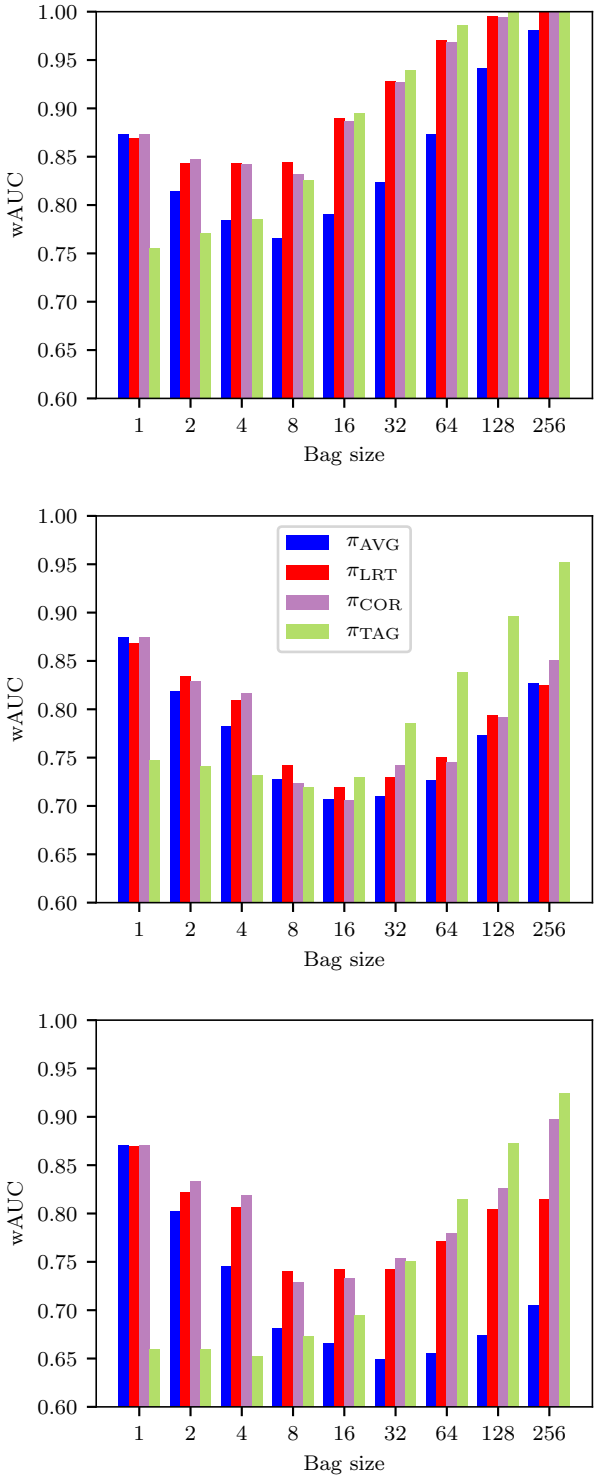


**Figure 5: Detection accuracy of Warden's SRNet2 in terms of wAUC versus the bag size for IMS, SLS, and MDS (top to bottom) with four different pooling strategies.**

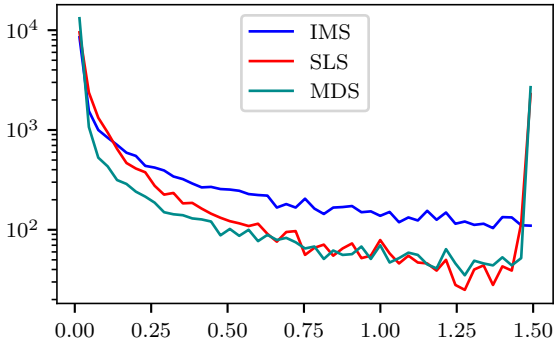


**Figure 6: Distribution of relative payloads across the training dataset for IMS, SLS, and MDS for bag size $B = 100$. Note that the new detector-informed senders are far more aggressive in assigning payloads to images with most images either being embedded with small payloads and a significant fraction embedded fully.**

Continuing our discussion of Figure 5, we now comment on which poolers are the most effective in detecting batch steganography across the same range of bag sizes and for all three senders. For large bags, the best detection is obtained with the tag-based detector across all three senders because it is trained on the closest stego source. The correlator $\pi_{\text{COR}}$ and the LRT $\pi_{\text{LRT}}$ typically provide similar performance and are significantly better than the simple average $\pi_{\text{AVG}}$. This difference is most striking for the MDS because the simple average is essentially a correlator with uniform payloads. Thus, the more non-uniform the payload distribution is the larger the difference (see, e.g., the performance of $\pi_{\text{AVG}}$ versus $\pi_{\text{COR}}$ across the senders).

The poor performance of the tag-based pooler for small bag sizes is understandable because, as a binary detector on stego images embedded with tags, it performs poorly (and is also more difficult to train) as less than 14% of images have payload larger than 0.05 bpp with a high number of images with extremely small payloads. It starts being effective only for larger bag sizes, which are more likely to contain almost fully embedded images.

In Figure 6, we display the histogram of payloads embedded in images from the training set for all three senders, $B = 100$, and $r = 0.3$ bpp. The SLS and MDS are clearly much more aggressive in using certain images close to their maximal embedding capacity than the IMS. This is because these senders are aware of the fact that the embedding is "invisible" to the sender's SRNet. Understandably, this leads to a large gain in security at least as long as the Warden uses the same type of single-image detector. If the Warden uses a different detector for pooled steganalysis, the almost fully embedded images may become detectable if their response curves are not as flat as the sender's. We take a look at this important aspect in Section 7.3.

Figure 6 also shows that MDS is slightly more aggressive than SLS in allocating very large or very small payloads. This can be understood from Eqs. (12) and (16) showing the payloads as functions of the RC slopes. The payload of the MDS is inversely proportional to the square of the slope, making this sender more agressive when

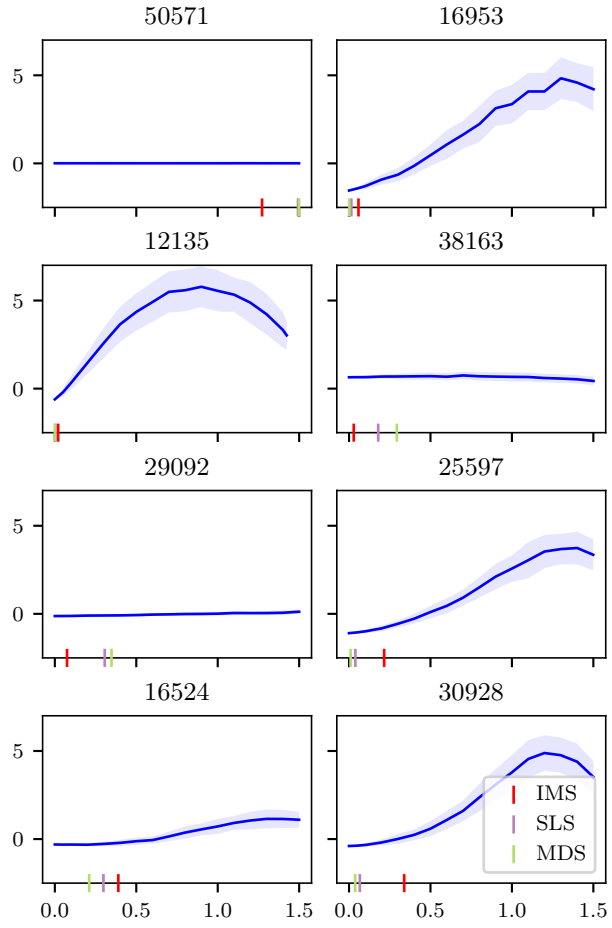


**Figure 7: SRNet's response curves for a bag of 8 images with payloads allocated to each image by IMS, SLS, and MDS marked on the $x$-axis.**

allocating the payload than the SLS. Figure 7 compares the three senders IMS, SLS, and MDS for a given bag of 8 images. For images 50571, 38163, and 29092, which have a flat response curve, the detector-aware senders embed larger payloads than the detector-agnostic IMS. For images with an increasing RC, such as 30928 and 25597, SLS and MDS are more conservative than IMS and allocate a smaller payload.

As the last experiment of this section, we include a study of the effect of the average communication rate $r$ on the optimal bag size. We limit our study to the SLS and SRNet2 as Warden's detector. Figure 8 shows wAUC of the best pooler as a function of the bag size for four rate $r$. Note that with increased rate the dip becomes shallower and also starts moving towards smaller bag sizes.

## 7.2 Effect of estimating the payloads

In any realistic scenario, the Warden may know the algorithms used to embed and spread the payloads but not Alice's data. All three senders compute the payload size to be embedded in each image from the cover image itself. The Warden, however, will need to estimate the payloads from the images at hand. The embedding

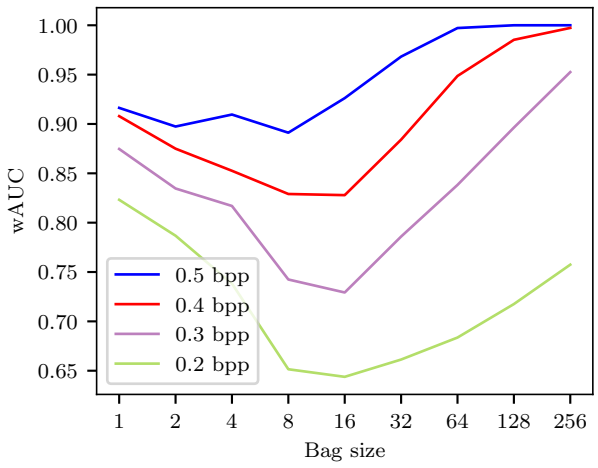


**Figure 8: Detection accuracy of the best pooler of SRNet2 for SLS versus the bag size $B$ and four communication rates $r$.**

changes themselves may skew the estimated payload size should the Warden estimate from a stego image. For the IMS, the effect of the embedding changes on computing the embedding costs (or Fisher information for model-based steganography) is known to be practically negligible [6, 20]. For the new detector-aware senders, however, the payloads are also determined from the cover response curves, which are more sensitive to the embedding itself. For an image that receives a large payload, the Warden may end up with a very different response curve. Thus, even if she knows the spreading strategy, the communication rate, and the type of the detector used by the senders, the payloads that potentially reside in the images will be subject to an estimation error and lower the detection accuracy. We study this effect in this section.

First, it is hard to imagine that it would be advantageous for the Warden to intentionally mismatch the payloads potentially embedded in the images. Thus, the Warden should estimate them using a detector that is as close to the senders' detector as possible. As our first experiment, in Table 1 we compare the accuracy of the pooled detectors for a Warden who trains

(1) SRNet2 on her dataset for $d^{\mathrm{W}}$ but uses the knowledge of the exact payloads $\alpha_{r,S}(\mathbf{X})$.
(2) SRNet2 on her dataset for $d^{\mathrm{W}}$ and uses SRNet2 for estimating the payloads from the images at hand $\alpha_{r,S}(\mathbf{Y})$.

Note that Case 1 corresponds to the setup assumed in the previous section. In Figure 9, we show the ROCs corresponding to two selected entries of Table 1. While estimating the payloads leads to a performance drop, the effect is minimal because most images in the bag hold small payloads and thus their response curves are close to the response curves of the corresponding covers. For images embedded with medium to large payloads, which however form a small portion of each bag, the estimated payloads may be very different. Figure 10 shows the the relative payloads used by the sender as determined from her version of SRNet1 versus payloads estimated using SRNet2 by the Warden from a HILL stego bag for SLS and MDS for $B = 16$ and $r = 0.3$ bpp.

| | $\pi_{\text{LRT}}$ | | $\pi_{\text{COR}}$ | |
|---|---|---|---|---|
| $B$ | 16 | 64 | 16 | 64 |
| SLS | .719 / .718 | .750 / .746 | .706 / .707 | .745 / .756 |
| MDS | .742 / .735 | .771 / .757 | .733 / .737 | .780 / .768 |

**Table 1: Accuracy (wAUC) of Warden's detectors for two senders, two bag sizes, and two pooling strategies with exact / estimated payloads. Warden's single-image detector is SRNet2, HILL 0.3 bpp.**

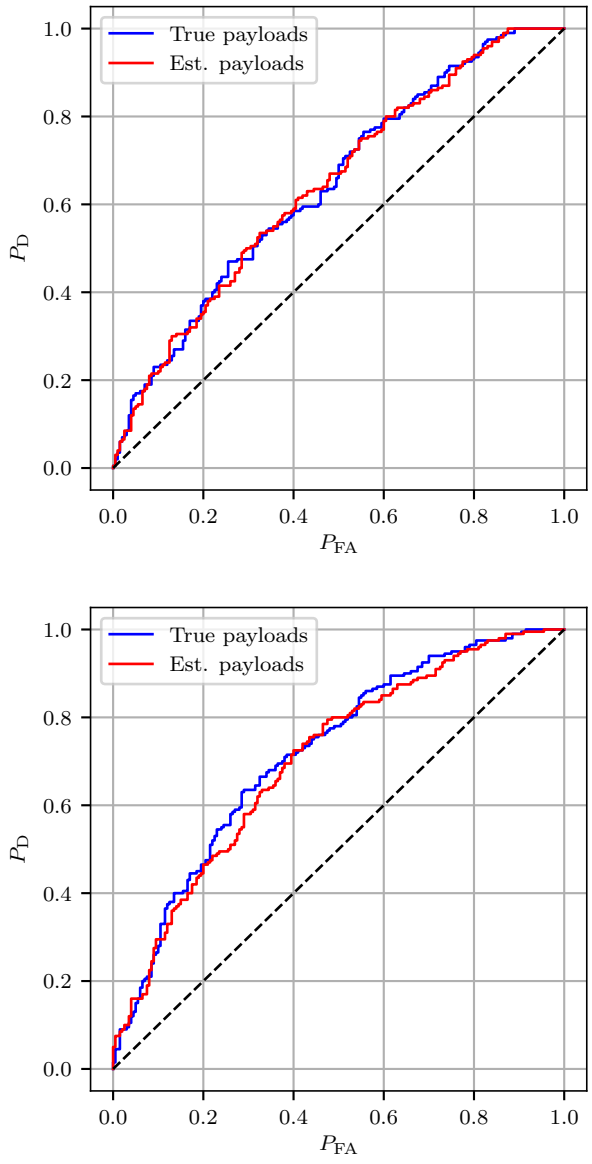


**Figure 9: ROCs of pooling strategies using Warden's SRNet2 with exact and estimated payloads; $\pi_{\text{LRT}}$ for SLS with bag size 16 (top) and $\pi_{\text{COR}}$ for MDS with bag size 64 (bottom) for HILL at $r = 0.3$ bpp.**

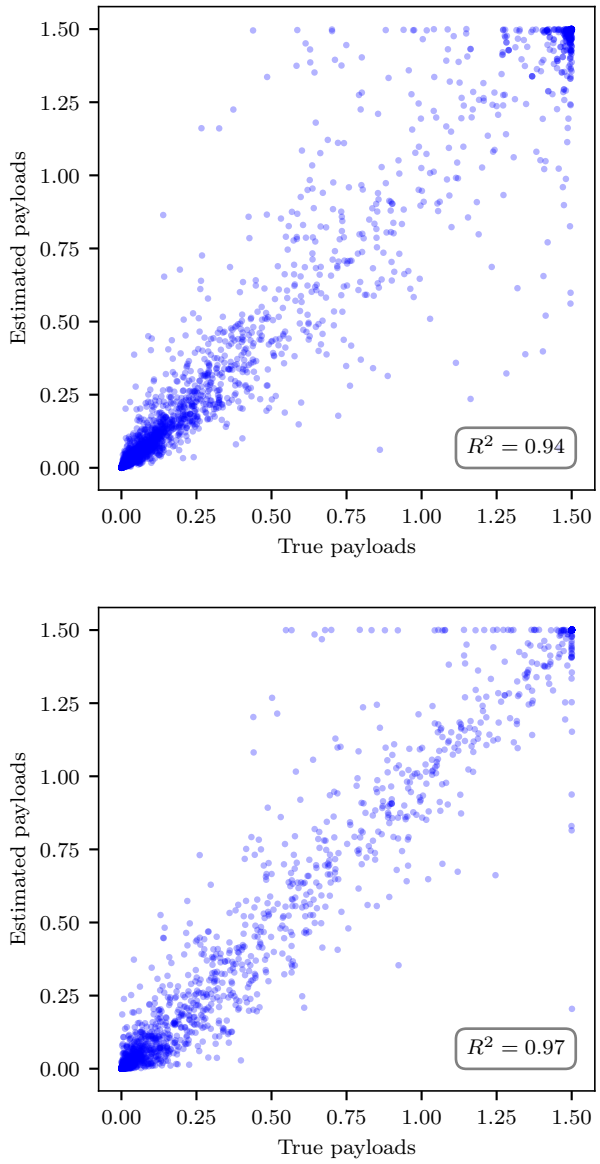


**Figure 10: Payloads estimated by the Warden using SRNet2 versus the true embedded payloads as determined by the senders using SRNet1 for the SLS (top) and MDS (bottom). Bag size 16, HILL, $r = 0.3$ bpp.**

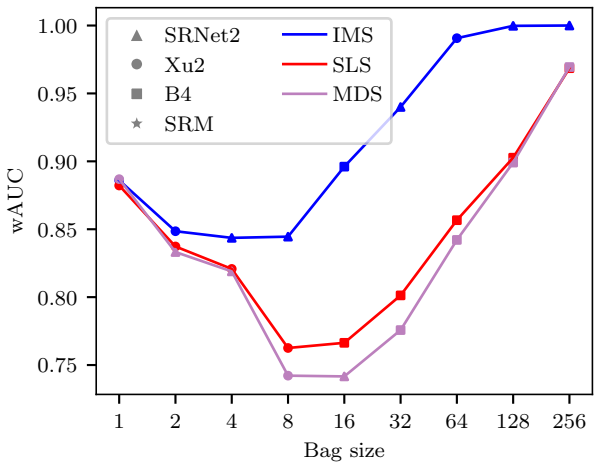


**Figure 11: Accuracy (wAUC) of the best detector and best pooling strategy versus the bag size for IMS, SLS, and MDS. The best detector for each setting is highlighted using a different marker.**

### 7.3 Devious Warden

Since the SLS and MDS use feedback from a detector, while being more powerful than IMS when the Warden uses the same type of detector for pooling, they could potentially become vulnerable when the Warden intentionally or unintentionally mismatches the single-image detector. In this section, we study such a devious Warden who trains a different architecture (or a completely different single-image detector) on her training set. Since the effect of using payloads estimated from the images at hand instead of exact ones is small, we give the Warden the exact same payloads for pooling. This has been adopted for simplicity due to excessive computational cost of having to estimate the average response curves. Moreover, it helps us isolate the effect of the mismatched detector for pooling. The experiments were carried out for the SLS, MDS and IMS with SRNet2, EfN B4, Xu2, and SRM for a range of bag sizes. The results displayed in Table (2) show that the Warden indeed may gain from mismatching the detector. The gain is, however, quite small, and the detector-aware senders still exhibit a much better security than the IMS. In Figure (11), we show wAUC of Warden's best detector from among 16 different possibilities (four pooling strategies and four single-image detectors) as a function of the bag size. The new spreading strategies perform significantly better than IMS, even when considering different CNN architectures, training sets, and a very different detector (SRM) than what Alice uses.

## 8 CONCLUSIONS

When communicating using steganography, the sender can be clever and choose to split the desired secret message among a bag of cover images to avoid being detected. In this paper, we determine the sizes of the payload chunks by inspecting how each image in the bag reacts to embedding in terms of changing the soft output of a steganography detector as a function of the payload size, the image's "response curve." Two such detector-informed senders are investigated for spatial-domain steganography: 1) a sender that makes sure that all images in the bag experience the same shift in the detector response and 2) a sender that minimizes the sum of squares of the shifts, which can be interpreted as a deflection coefficient for a binary test distinguishing stego images from covers naturally corrupted by acquisition noise.

Using feedback from a detector indeed brings substantial improvement over the previously proposed image-merging sender that considers the bag as a single large image. The detectability as a function of the bag size for a fixed secret communication rate initially decreases, because the sender makes better use of all available covers, and then starts increasing due to the square root law since a fixed rate is maintained. We experimentally determined that the optimal bag size is 8–16 images per bag depending on the average communication rate.

On the detection side, we study three different strategies for the Warden to pool the outputs of her single-image detector: 1) correlator of the outputs with the expected detector output increase, 2) likelihood ratio test based on actual models of the detector output, and 3) detector trained on payload tags that the images would receive for sufficiently large bags. The likelihood ratio was the best pooling strategy for small to moderate bag sizes up to 16 while the tag based detector performed better for bag sizes larger than 16.

Using feedback from a detector for spreading can potentially backfire as the Warden may use a different detector for pooling. We looked into this issue in great detail by training alternative deep learning architectures as well as older rich-model based detectors. We discovered that doing so increases the Warden's accuracy, but not substantially and the detector-aware senders are still much more secure than the IMS.

In the future, we intend to further investigate the problem of optimal bag size by modeling the statistical collection of response curves. We also intend to explore the JPEG domain.

## ACKNOWLEDGMENTS

The work on this paper was supported by NSF grant No. 2028119.

## REFERENCES

[1] P. Bas. Steganography via cover-source switching. In *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, December 4-7 2016.

[2] M. Boroumand, M. Chen, and J. Fridrich. Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 14(5):1181–1193, May 2019.

[3] J. Butora, Y. Yousfi, and J. Fridrich. How to pretrain for steganalysis. In D. Borghys and P. Bas, editors, *The 9th ACM Workshop on Information Hiding and Multimedia Security*, Brussels, Belgium, 2021. ACM Press.

[4] R. Cogranne and J. Fridrich. Modeling and extending the ensemble classifier for steganalysis of digital images using hypothesis testing theory. *IEEE Transactions on Information Forensics and Security*, 10(2):2627–2642, December 2015.

[5] R. Cogranne, Q. Giboulot, and P. Bas. ALASKA–2: Challenging academic research on steganalysis with realistic images. In *IEEE International Workshop on Information Forensics and Security*, New York, NY, December 6–11, 2020.

[6] 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.

[7] J. Fridrich, M. Goljan, D. Soukal, and P. Lisoněk. Writing on wet paper. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VII*, volume 5681, pages 328–340, San Jose, CA, January 16–20, 2005.

[8] 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.

[9] V. Holub, J. Fridrich, and T. Denemark. Universal distortion design for steganography in an arbitrary domain. *EURASIP Journal on Information Security, Special*

| $B$ / $\pi$ | 16 / $\pi_{\text{LRT}}$ | | | | 64 / $\pi_{\text{TAG}}$ | | | |
|---|---|---|---|---|---|---|---|---|
| | SRNet2 | Xu2 | B4 | SRM | SRNet2 | Xu2 | B4 | SRM |
| SLS | .7190 | **.7324** | .7209 | .6799 | .8382 | .7973 | **.8567** | .7127 |
| MDS | **.7416** | .7259 | .7393 | .7030 | .8150 | .7806 | **.8421** | .7166 |
| IMS | **.8902** | .8836 | .8877 | .6664 | .9858 | **.9907** | .9842 | .7244 |

**Table 2: Accuracy (wAUC) of Warden's detectors for three senders, two bag sizes, with two pooling strategies for HILL 0.3 bpp.**


*Issue on Revised Selected Papers of the 1st ACM IH and MMS Workshop*, 2014:1, 2014.

[10] X. Hu, J. Ni, W. Zhang, and J. Huang. Efficient JPEG batch steganography using intrinsic energy of image contents. *IEEE Transactions on Information Forensics and Security*, 16:4544–4558, 2021.

[11] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, volume II. Upper Saddle River, NJ: Prentice Hall, 1998.

[12] A. D. Ker. Batch steganography and pooled steganalysis. In J. L. Camenisch, C. S. Collberg, N. F. Johnson, and P. Sallee, editors, *Information Hiding, 8th International Workshop*, volume 4437 of Lecture Notes in Computer Science, pages 265–281, Alexandria, VA, July 10–12, 2006. Springer-Verlag, New York.

[13] A. D. Ker. Batch steganography and the threshold game. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, pages 04 1–13, San Jose, CA, January 29–February 1, 2007.

[14] A. D. Ker. Perturbation hiding and the batch steganography problem. In K. Solanki, K. Sullivan, and U. Madhow, editors, *Information Hiding, 10th International Workshop*, volume 5284 of Lecture Notes in Computer Science, pages 45–59, Santa Barbara, CA, June 19–21, 2008. Springer-Verlag, New York.

[15] A. D. Ker. The square root law of steganography. In M. Stamm, M. Kirchner, and S. Voloshynovskiy, editors, *The 5th ACM Workshop on Information Hiding and Multimedia Security*, Philadelphia, PA, June 20–22, 2017. ACM Press.

[16] A. D. Ker and T. Pevný. The steganographer is the outlier: Realistic large-scale steganalysis. *IEEE Transactions on Information Forensics and Security*, 9(9):1424–1435, September 2014.

[17] A. D. Ker and Tomas Pevný. Batch steganography in the real world. In J. Dittmann, S. Craver, and S. Katzenbeisser, editors, *Proceedings of the 14th ACM Multimedia & Security Workshop*, pages 1–10, Coventry, UK, September 6–7, 2012.

[18] L. Li, W. Zhang, C. Qin, K. Chen, W. Zhou, and N. Yu. Adversarial batch image steganography against CNN-based pooled steganalysis. *Signal Processing*, 181:107920–107920, 2021.

[19] V. Sedighi, R. Cogranne, and J. Fridrich. Practical strategies for content-adaptive batch steganography and pooled steganalysis. In *Proceedings IEEE, International Conference on Acoustics, Speech, and Signal Processing*, March 5–9, 2017.

[20] V. Sedighi and J. Fridrich. Effect of imprecise knowledge of the selection channel on steganalysis. In J. Fridrich, P. Comesana, and A. Alattar, editors, *3rd ACM IH&MMSec. Workshop*, Portland, Oregon, June 17–19, 2015.

[21] M. Sharifzadeh, M. Aloraini, and D. Schonfeld. Adaptive batch size image merging steganography and quantized Gaussian image steganography. *IEEE Transactions on Information Forensics and Security*, 15:867–879, 2020.

[22] 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*, 14(8):2074–2087, 2019.

[23] Y. Yousfi, J. Butora, E. Khvedchenya, and J. Fridrich. ImageNet pre-trained CNNs for JPEG steganalysis. In *IEEE International Workshop on Information Forensics and Security*, New York, NY, December 6–11, 2020.