Cover-Source Mismatch in Steganalysis: Systematic Review

Antoine Mallet

antoine.mallet@utt.fr

Universite de Technologie de Troyes https://orcid.org/0009-0006-0479-564X

Martin Benes

University of Innsbruck: Universitat Innsbruck

Rémi Cogranne

Universite de Technologie de Troyes

Research Article

Keywords:

Posted Date: January 15th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3812991/v1

Version of Record: A version of this preprint was published at EURASIP Journal on Information Security on August 12th, 2024. See the published version at https://doi.org/10.1186/s13635-024-00171-6.

Cover-source Mismatch in Steganalysis: Systematic Review

Antoine Mallet[1], Martin Beneš[2], and Rémi Cogranne[1]

[1]Univerité de Techonologie de Troyes, Troyes, France
[2]Univerity of Innsbruck, Innsbruck, Austria

January 3, 2024

Abstract

Operational steganalysis contends with a major problem referred to as the cover-source mismatch (CSM), which is essentially a change of distribution caused by different parameters and settings over training and test data. Despite it being of fundamental importance in operational context, the CSM problem is often overlooked in the literature. With the goal to increase the visibility of this problem and attract the interest of the community, the present paper proposes a systematic review of the literature. It summarizes gathered knowledge and major open questions over the last 20 years of active research on CSM: terminology, methods of measurement, known causes, and mitigation strategies. Over 100 papers exploring, mitigating, assessing or discussing steganalysis under train-test mismatch were collected by sampling scholar databases, and tracing references, cited and generated. For image steganalysis, the literature provided enough evidence to quantify the impact of causes, and the effectiveness of mitigation strategies.

Declaration *Availability of data and materials.* The list of bibliographical reeferences is given at the end of the paper. Additionally, a summary of the papers, as well as analytical datas will be given upon acceptance of the paper. *Competing interests.* Authors declare no competing interests. *Funding.* This work is funded by the EU's Horizon 2020 programme under the grant agreement No. 101021687(UNCOVER), and the french ANR PACeS project No. ANR-21-CE39-0002. *Author's contributions* The bibliographical research and most of the writing were conducted by Antoine Mallet and Martin Beneš. Rémi Cogranne provided crucial advices and complementary writing. *Acknowledgments* Additionally, authors would like to thank their respective supervisors.

1 Introduction

Steganography is often referred to as the art and techniques of cover communication. It aims at secretly exchanging a sensitive information by hiding it into a so-called cover object. This creates a stego-object which, in order to preserve the furtiveness of the secret communication, should look as inconspicuous as possible.
To ease this process, the cover object ought to be commonly encountered, in order not to raise suspicion. It should be easy to modify and carry enough entropy to accommodate the secret message; good examples of such suitable cover objects include digital media – images, audio, or video – texts, computer network packets or even program executable codes [KP16; Fri09].

As for any secured communication, the stego-object is sent over an insecure channel which, in the worst case, is assumed monitored or controlled by an adversary referred to as the steganalyst. Unlike the other scenario of communication security, the steganalyst aims, in the very first place, at detecting the presence of a hidden secret message either by thorough statistical analysis or by searching for trademarks or "signatures" of a specific technique.

Steganography and steganalysis, thus, constitute a game of cat and mouse. In academic studies, the Kerckhoffs' principle[1] is often advocated in order to justify that the steganalysis is carried out with knowledge on all necessary properties of the inspected objects. This also includes the potential embedding method as well as access to large representative datasets [PF07; LK11]. Of course, steganography has been developed in this setting, which is the most stringent.

On the opposite, in a real-world operational context, the steganographer and the steganalyst only have very limited access to each other's information. The

[1] The Kerckhoffs' principle essentially states that security must always rely solely on the key and that the rest of the communication system and its settings must be publicly known.

steganalyst selects a detector, and the steganographer picks the steganographic strategy and the cover. The exact original cover object is unknown to the steganalyst, but it is generated from a noisy process which is defined as the *cover source*. As advocated in [KP14b] the steganalyst can hardly know this *cover source*; it can only, at best, be estimated with an accuracy that depends on the nature and number of objects provided the steganographer.

From this scenario naturally raises the problem, for the steganalyst, of "designing" or "training" a detector on a *cover source* that differs from the one used by the steganographer: this is referred to as the *cover source mismatch*.

In the broad field of statistical learning, this phenomenon is known as the *distribution shift* and it occurs when the statistical properties of training and testing data differ. The main symptom is a deterioration of the model performance in the production environment. Distribution shift occurs in all possible application of statistics and machine learning, such as to cite a few, medical imaging [GL22], computer vision [Pen+19], reinforcement learning [ZQW20], natural language processing [BDP07] and speech recognition [Gon95].

However, it should be noted that the aforementioned tasks operate mostly on a semantic level. In other words, while the acquisition and processing setting do matter, it has a limited impact as compared to the presence of the pattern of interest. In addition, it is often possible to adjust the training for a specific source.

The peculiarity of steganography, and the related field of digital forensics [LK19], is that the signal of interest is extremely weak while the cover-source mismatch has a much stronger impact; this often yields catastrophic performance drops, which make the steganalysis merely ineffective. In [Gib+20], using different capturing devices increased the error rate from 15% to random guessing. According to [KP16], "CSM is one of the main factors negatively affecting the deployment of steganalysis in the real world."

The present paper presents a systematic review of literature on CSM in steganalysis, and mainly aims at addressing the following research questions:

Research Question 1: What are the studied causes of CSM in the literature ?

We list the known causes of CSM studied in the literature, and look into the research trend over time.

Research Question 2: How impactful are the known causes of CSM ?

We quantitatively assess the impact of CSM causes identified by answers to RQ. 1.

Research Question 3: What are the existing mitigation strategies against CSM ?

Similar to RQ. 2, we assess the relative effectiveness of the existing mitigation strategies.

At the time of writing of this survey, RQ 2 and RQ 3 can only be meaningfully answered for images media, which constituted nearly all the literature sampled using the strategy from Section 5.1. While the present paper aims at encompassing all forms of steganalysis, it only assumes digital images in Sections 6 and 7.

The rest of this survey is organized as follows: Section 2 presents prior work on CSM in steganalysis and other applications of machine learning. Section 3 describes the train-test mismatch problem in steganalysis and its implications. Section 4 presents the methods of measuring the CSM. Section 5 explains the bibliometric methodology for literature collection and result aggregation. Section 6 surveys identified causes of CSM, and quantifies their impact. Section 7 reviews strategies to mitigate CSM and quantifies their impact. Section 8 discusses the findings, answers the research questions, and poses still open questions, and Section 9 concludes.

2 Related Works

The symptoms of the CSM problem have been identified for a little less than 20 years, as image steganalysis was in its infancy. It has been empirically observed that steganalysis techniques perform differently over different datasets [KSM05; Can+08]. The significance of the CSM impact has been clearly acknowledged for the first time in 2010 during the BOSS open contest (Break Our Steganographic System) as the organizers added in the testing set data generated with a different source[2]. This period also coincided with the increasing use of machine learning in steganalysis, which requires a training phase after which the classifier can be used and evaluated on a different testing set. While this problem was unanimously recognized as an important deadlock for practical applications of steganalysis [Ker+13], it has been only seldom studied. In particular, it was only in 2018 that the source of CSM was thoroughly studied [GCB18; BBB18; Gib+20] by a comprehensive evaluation of the contribution of each step of a generic image processing pipeline (IPP) to CSM.

[2] More precisely, the training set of BOSS was made of raw images processed with the very same script. On the opposite, the testing set included out-of-camera jpeg compressed image. An important drop of performance of all competitors was observers on this testing subset.

In the meantime, several strategies have been suggested to mitigate the impact of CSM on steganalysis performance [BCE10; LK12; PBC14; KP14a; LM16b]. However, despite its severity on modern steganalysis, the problem of CSM remains largely unexplored and existing steganalysis survey papers discuss CSM only briefly [RRG19; Hus+20]. To the authors' best knowledge, the present paper is the first systematic review of the literature on CSM.

2.1 Comparison with other fields of ML

Essentially, the CSM is a form of data heterogeneity, which is a common problem in many fields of application of ML. While CSM hasn't been the focus of any survey paper yet, one can find such surveys in other fields. We selected surveys focusing on data heterogeneity from 5 fields: medical imaging (MI), mechanical components monitoring (MCM), speech recognition (SR), natural language processing (NLP), and temporal reasoning (TR).

We provide a summary – with regard to the proposed RQs, of our readings in Tab. 1. Regarding RQ. 1, we see that each field has recognized its own difficulties. A striking observation is that some are conceptually very close. Patients in MI, components in MCM, authors in NLP can be understood as similar causes of mismatches; the same goes for working conditions in MCM, context in NLP and environment in SR. On the other hand, languages are conceptually rather specific to the fields of NLP. Mismatch from the acquisition devices and their calibration is expected to occur in every field, but is foremost studied in MI.

Similar comments can be made for RQ. 3. With the advances in ML, general frameworks to deal with mismatch have been proposed, domain adaptation (DA) being among the most popular. It provides tools like shallow & deep feature matching ML (e.g. TCA, CORAL). Domain adaptation is particularly useful when labels are expensive, such as in MI, MCM based on operational data, or some NLP tasks, as it works in semi- and unsupervised setups. When labels are available in large quantities, however, pre-training and transfer learning can be very effective. There also exist specific solutions, depending on the problem and type of data. SR and MCM are basically signal processing tasks: they benefit from their own solution.

Steganalysis faces similar issues (e.g. acquisition device and content), and has similar solutions (domain adaptation and transfer learning). The peculiarity of steganalysis is that it operates at roughly the same level as the impact of the causes, although the impact of steganography is actually much lower than the impact of the CSM. On another note, as we will cover in Sec. 6, the diversity of the causes of CSM seems much greater than in the other fields.

3 Background on Mismatches

This section introduces different types of mismatch, mismatch in cover sources (Sec. 3.1), mismatch in steganography (Sec. 3.2), and mismatch in the context of pool steganalysis (Sec. 3.3).

3.1 Mismatch of Cover Sources

Steganography is carried out in two main phases: first the steganographer generates a cover from a *cover source*, that is a non-deterministic acquisition process followed by a processing pipeline. Both the acquisition and the processing pipeline consist of several steps which can be highly parametrized. Then, this cover object is used as an input for a steganographic algorithm, also characterized by a set of parameters, to hide a secret message into the so-called stego-object.
Potentially, a problem of training-testing set mismatch can occur when changing any parameter of any step, from acquisition of the cover objects up to the generation of the stego-object. In practice, however, two objects are never generated in exactly the same manner. In addition, different changes yield different impacts, with more or less important effects.

Adopting the same practical point of view that is used is almost all prior works, we shall define these concepts in relation with its use in steganography and its impact on steganalysis.

Definition 1 [Cover source] A cover source is entirely defined by the steps both the acquisition and processing pipeline are made of, the order these steps and the parameters used therein.
As a consequence, a cover source is a noisy process producing covers whose statistical characteristics are identical.

While this definition of the cover source is rather straightforward, it is hardly related with practical uses and applications. Even the second part, the corollary that objects from the same cover source share the same properties, remains rather impractical; indeed, it is still unclear which property has a significant impact on the usage of steganography and/or steganalysis. From a practical point of view, the cover source has little or no impact on steganography, which very often operates over each object independently. On the opposite, steganalysis generally exploits a detector that is trained and evaluated over a set of objects.

Definition 2 [Cover-source mismatch] Cover-source mismatch (CSM) occurs in steganalysis when using, for designing a detector, a training dataset whose

ML Field	Surveys	Causes of mismatch	Mitigation strategies
MI	[GL21]	Difference of acquisition device, parameters, and patients between datasets.	Domain Adaptation to cope with the low number of available labels.
MCM	[Yao+23]	Different working conditions. Different monitored components. Transfer from simulations to the real-world.	Transfer Learning; Domain Adaptation, e.g. shallow and deep feature matching, GANs,...
SR	[Zha+18]	Noisy environments	Preprocessing, features extraction. Using acoustic/language models.
NLP	[RP20]	Cross-lingual learning. Writers, contexts and dates also cause mismatches.	Data-centric (pre-training, pseudo-labelling) & model-centric (adversarial networks, autoencoders) DA.
TR	[ZH07], [Luo+17]	Difference in the datasets' origins, e.g. time granularity, or difference between temporal textual expressions.	Temporal data management, e.g. data integration, or NLP-based methods.

Table 1: Summary of the answers to the proposed RQs in other fields of application of ML: medical imaging (MI), mechanical component monitoring (MCM), speech recognition (SR), natural language processing (NLP), temporal reasoning (TR).

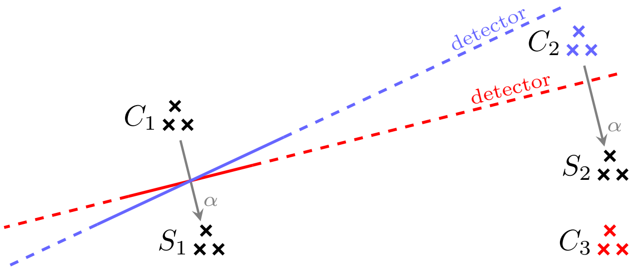


Figure 1: Illustration of cover-source mismatch for detectors distinguishing C_1 and S_1.

statistical properties differ from those found in the target testing dataset.

As stated above, the cover source mismatch affects the steganalyst. However the previous definitions are hardly applicable in practice. Therefore, we propose below a practical definition for the problem of the cover-source mismatch exploiting the impact it has for operational steganalysis:

Definition 3 [Cover-source mismatch problem] The cover-source mismatch problem is the ensuing degradation of steganalyser performance resulting from the cover-source mismatch.
Therefore it is essentially observed as the loss of performance when a dataset from a different cover source is used for training.

Def. 3 of the CSM problem explicitly addresses the problem of measuring the impact on performance of steganalysis because this is how CSM is actually observed. This aspect is addressed in more detail in Sec. 4. Before that, the rest of the present section describes how CSM affects the training of a detector as well as the different types of CSM.

Fig. 1 depicts an illustrative scenario of three cover sources, C_1, C_2, and C_3 and exemplifies the preceding definitions. The geometric proximity corresponds to cover source similarity: C_2 and C_3 are more similar to each other than to C_1. Note that the figure also illustrates that the embedding shift is often smaller and similar in direction [3] as compared to the wide diversity between cover sources, hence the critical impact that CSM may have. These elements illustrate a typical configuration of cover-source mismatch.

Additionally, a given steganographic embedding at rate α is performed, which shifts C_1 to stego S_1 and C_2 to stego S_2. Two steganalysis detectors, shown in blue and red respectively, were then trained to distinguish between the same cover source C_1 and corresponding stego objects S_1. The difference between the two can come from the selected model, hyperparameters, and initial weights. Both detectors discriminate well C_1 from S_1. However, the one illustrated in red is not subject to the CSM problem when applied on cover source C_2. On the opposite, both detectors perform poorly over C_3 because of the CSM problem: cover objects erroneously belong to decision region "stego" resulting in a very high false-positive detection rate (also referred to as type I error): this is the cover-source mismatch problem.

We believe that Fig. 1, along with Def. 2, illustrate an important – and often overlooked – fact: the CSM problem, as a degradation of a detector's performance, depends on two things: the CSM configuration itself, but also the settings of the detector. We further discuss ways of measuring CSM in Sec. 4.1 and 4.2.

The reader might also wonder how to generalize better to a very large amount of cover sources, possible unseen. Mitigation strategies and their assessment is

[3] This is usually referred to as the shift hypothesis [Ker06; CSF17].

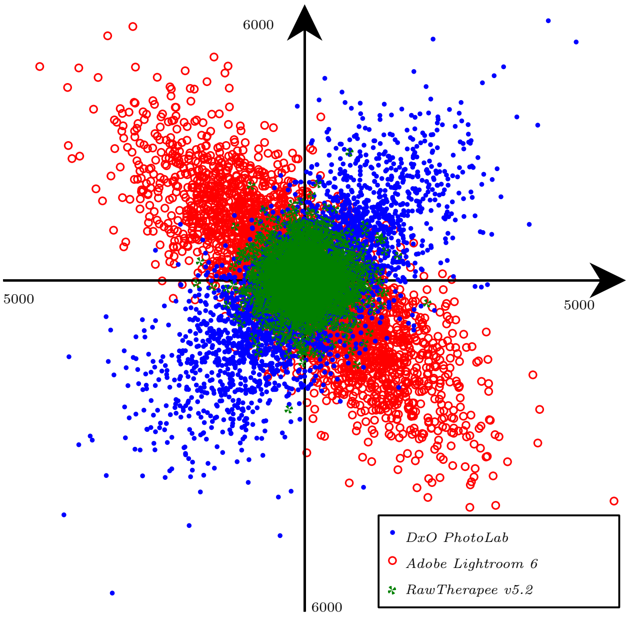


Figure 2: Scatter plots that show the empirical joint distribution of unquantized DCT coefficients $(0, 7)$ and $(7, 0)$ for images corrupted with only i.i.d Gaussian noise and then developed with three different software.

addressed in detail in Sec. 7.

Let us conclude this definition section by a real-world example shown in Fig. 2 and 4 taken from [Gib+20]. First, Fig. 2 show scatter plot of co-occurrence (i.e. joint empirical distribution) between two randomly selected DCT coefficients resulting from JPEG cover sources. The images from the exact same sources with the exception that they are converted from RAW files to uncompressed TIFF images using different software. More precisely, we note that the simplest development was used (essentially made of demosaicing, white balance and gamma-correction).
It can be clearly seen from Fig. 2 that the correlation between the selected DCT coefficients are almost in complete opposition. In addition, one can note that the variance of marginal distribution also changes significantly, especially for *RawTherapee*, see green dots on Fig. 2.

Last, Fig. 4 exemplifies how the CSM can have a dramatic impact on detector performance. Depending on the detector and the sources, the CSM problem can be either barely noticeable or make a detector no better than a random-guesser.

3.2 Mismatch in Steganography

Detectors must also face heterogeneity in the steganography in training and test sets. This is a separate type of mismatch than CSM, because steganographer choice of cover sources and steganography are independent. In addition, as we shall explain in Sec. 4.2, measuring the CSM problem via features inherently separates the two. However, both mismatches exhibit the same symptoms, and existing mitigation strategies against CSM can be applied to both.

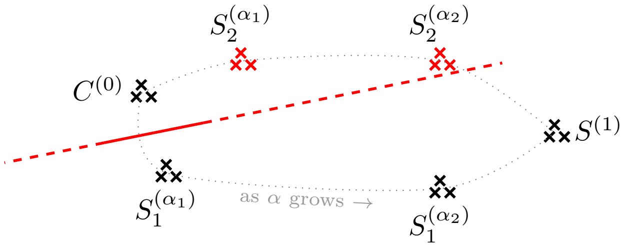


Figure 3: Stego-source mismatch between schemas S_1 and S_2, and embedding rates $\alpha_1 < \alpha_2 < 1$.

The literature uses either "stego-source mismatch" (SSM) [DBF16; Lea+22], and "stego-algorithm mismatch" (SAM) [Rei+19]. We suggest a new name, stego-scheme mismatch (SSM), because it is more general and hence encompasses the few possible causes of the mismatch of this kind.

Definition 4 [Stego-scheme mismatch] Stego-scheme mismatch (SSM) is a mismatch of the distributions of stego objects stemming from a given cover source. It is caused either by using different stego-schemes or different settings to embed data in cover objects.

SSM has two known causes: stego embedding, and embedding rate α, illustrated in Fig. 3 by orientation and length of the stego shift. Embedding into a cover source $C^{(0)}$ with two schemes S_1 and S_2 and payload $\alpha < \alpha_{\max}$ causes different stego shifts. The dotted trajectories of both schemes suggest that for $\alpha = \alpha_{\max}$ the embedding is equivalent to LSB matching, denoted $S^{(1)}$.[4]

Note that the shift hypothesis [Ker06; CSF17] states that, roughly speaking, the trajectory are the same regardless of the cover sources. The same embedding causes the same impact (the same "shift") on the stego distribution.

SSM is more important for test sets with (1) lower α, or (2) less detectable embedding scheme [Che+17].
A *universal* detector is capable to generalize to any stego method.

[4] Note that in practice the maximal achievable payload $\alpha_{\max}$ depends on the coding method. With LSB matching, ternary embedding allows embedding up to $log_2(3) \approx 1.585$ bit per element.

3.3 Mismatch in pool steganalysis

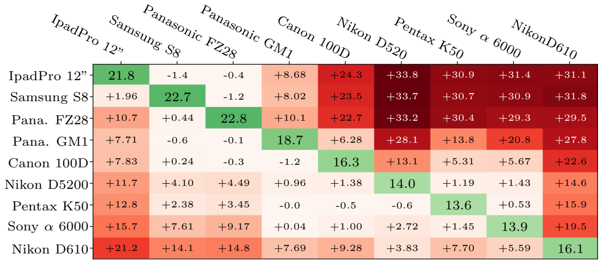
Pooled steganalysis is the problem that arises when one tries to inspect a group of several objects over which one sole decision must be made (typically "all these objects are covers" vs. "some objects are stego-files"). Pooled steganalysis either indirectly addresses CSM or is affected by it.

On the one hand, several works addressed the problem of pooled steganalysis when it aimed at identifying guilty actors, in which case this framework can lift the problem of CSM provided that "*each actor used a source of cover objects, different from sources used by other actors*" [KP12a]. However, this means that the steganalyst needs to know the source of each actor which is hardly possible in the real-life, "*unless the suspected steganographer is considerate enough to supply [...] the cover source*" [KP14b].
Under the same assumption, a similar approach was adopted in [LM23] by detecting inconsistencies in the classifier in order to identify the suspect(s).

Another look at the mismatch was studied from pooled steganalysis when the strategy of the steganographer is not known: that is how the payload is spread over several objects. In this case, the steganalyst typically faces a stego scheme mismatch as the embedding rate for each object is unknown.
A robust statistical sequential method based on CUSMU (Cumulative Sum) was proposed in [Cog15] while the method developed in [PN15] is based on the histogram of the classifier "soft-output" (before thresholding). The problem was also leveraged in [CSF17; KDF23]: in an adversarial setting, it has been proposed to spread the payload in order to create the hardest stego scheme mismatch hence reducing detectability.

4 Measuring the CSM Problem

In Def. 2 and 3 we have taken particular care to separate the CSM from the problem it generates through its impact on steganalysis. Thus, measuring severity is critical for the very definition of the CSM problem but also in order to study its causes (RQ. 2) and for successful mitigation of CSM (RQ. 3). In this section, we describe the two main approaches that have been used for assessment of the CSM problem. First, and most obvious, in Sec. 4.1, mismatch problem is measured via performance of a detector. Second, Sec. 4.2 describes the studies measuring CSM problem using a distance in features spaces.



	IpadPro 12"	Samsung S8	Panasonic FZ28	Panasonic GM1	Canon 100D	Nikon D520	Pentax K50	Sony α 6000	NikonD610
IpadPro 12"	21.8	-1.4	-0.4	+8.68	+24.3	+33.8	+30.9	+31.4	+31.1
Samsung S8	+1.96	22.7	-1.2	+8.02	+23.5	+33.7	+30.7	+30.9	+31.8
Pana. FZ28	+10.7	+0.44	22.8	+10.1	+22.7	+33.2	+30.4	+29.3	+29.5
Pana. GM1	+7.71	-0.6	-0.1	18.7	+6.28	+28.1	+13.8	+20.8	+27.8
Canon 100D	+7.83	+0.24	-0.3	-1.2	16.3	+13.1	+5.31	+5.67	+22.6
Nikon D5200	+11.7	+4.10	+4.49	+0.96	+1.38	14.0	+1.19	+1.43	+14.6
Pentax K50	+12.8	+2.38	+3.45	-0.0	-0.5	-0.6	13.6	+0.53	+15.9
Sony α 6000	+15.7	+7.61	+9.17	+0.04	+1.00	+2.72	+1.45	13.9	+19.5
Nikon D610	+21.2	+14.1	+14.8	+7.69	+9.28	+3.83	+7.70	+5.59	16.1

Figure 4: Results from [Gib+20] depicting steganalysis error rate for different camera models at the lowest ISO sensitivity.

4.1 Measuring a mismatch via detectors

One way to measure the CSM is through comparison of the detector error ϵ in matched and mismatched scenarios. Within the framework of modern steganalysis, based on machine learning, the clairvoyant scenario is the one in which training datasets perfectly match the testing datasets over which the detection performance is measured.[5] This case, without CSM, represents the "ideal" baseline error of steganalysis and is referred to as the *intrinsic difficulty* of a given source (see Def. 5).

The mismatched scenario error is the one when the training is performed over a dataset which presents some statistical differences with the testing dataset. This corresponds to Def. 2 of a CSM scenario. The ensuing degradation of detector performance is referred to as the inconsistency between sources, see Def. 6.

In order to illustrate as clearly as possible these concepts, Tab. 2 shows a toy-example with two sources. Note that, just alike in Fig. 4, rows represents the dataset used for training the detector while columns are for the testing dataset, on which detection performance is actually measured. As labelled in Tab. 2, the detection error-rate on the diagonal measures the intrinsic difficulty as in those cases training and testing datasets match. On the opposite, the out-of-diagonal results report the detection error-rate in case of mismatches hence the degradation error-rate due to the mismatches. In the rest of the present section, we will use the following notations: ϵ_{XX} represent the detection error-rate when training on the dataset X and testing on the same dataset X. In this case without a mismatch, the intrinsic difficulty is measured and is sometimes denoted ϵ_X for short. On the opposite,

[5] Note that some steganalysis methods do not use machine learning. However, these statistical detection methods [TRC14; CF15] generally include some parameters (weights, detection threshold, etc. ...) whose settings require some cover and stego examples hence may be subjuct to the CSM problem [KSM05; Can+08].

ϵ_{YX} represents the case when training the detector on the source Y while testing on the source X (since training is carried out first, it is the first variable in the notation we adopt). In this case of CSM, the detection performance reveals the CSM problem.
Eventually, note that in machine learning the detection performance is usually measures with the overall accuracy. On the opposite, in steganalysis the detection error-rate is generally the total probability of error (under equal prior), denoted P_E, without distinguishing false positive from false negative. Therefore in the following we will often refer to the error-rate as a measure of detection performance (the lower error-rate, the high performance).

With these notations, we can now formally define the notions that have been adopted in the literature for measuring CSM using detector error-rate:

Definition 5 [Intrinsic difficulty] The intrinsic difficulty ϵ_{XX} or ϵ_X of a source X is the error-rate of a detector that is trained and tested on X. It expresses how difficult it is for a detector to detect a steganographic scheme of embedding rate α in covers from a cover source X.

The *intrinsic difficulty* represents the error-rate of steganalysis on a given source without any mismatch. Therefore, this criterion is not related to the CSM problem, but it must be taken into account, serving as a baseline, to quantify the CSM problem.
Note that the intrinsic difficulty can vary greatly, even for similar sources: Fig. 4 show that the camera model can impact by almost 10% the error-rate.

Definition 6 [Inconsistency] The inconsistency ϵ_{XY} is the error-rate of a detector, trained on source X and evaluated on source Y.

The source inconsistency alone carries little information by itself, and should be accompanied by the corresponding intrinsic difficulty. This reference can be made more explicit by reporting the regret [ŠAP22; Abe+22; Abe+23]:

Definition 7 [Regret] The regret $R_{XY,Y}$ is the difference between the inconsistency ϵ_{XY} and the intrinsic difficulty ϵ_Y, as shown in Eq. 1,

$$R_{XY,Y} = \epsilon_{XY} - \epsilon_Y. \tag{1}$$

To exemplify this concept, one can have a look back at Tab. 2. The inconsistency is higher when testing on the source Y, $\epsilon_{X,Y} = 0.38$ than when testing on the source X, $\epsilon_{Y,X} = 0.33$. However, the intrinsic difficulty is also much higher on the source Y, $\epsilon_Y = 0.35$, than on the source X, $\epsilon_X = 0.2$.
Therefore, the "*relative degradation due to the CSM problem*", which is defined as the regret by Def. 7, is actually much lower when testing on the source Y, $R_{XY,Y} = 0.03$ as contrast to when testing on the source X, $R_{YX,X} = 0.13$.

		Test on source X	 source Y
Train on	source X	0.2 ← ϵ_{XX}	0.38 ← ϵ_{XY}
	source Y	0.33 ← ϵ_{YX}	0.35 ← ϵ_{YY}

Intrinsic difficulty

Inconsistency

Table 2: An illustrative example of presenting mismatched scenario results measured via a detector error-rate. The diagonal contains intrinsic difficulties, off-diagonal values are the inconsistencies. Inconsistencies can be replaced by the difference with the corresponding diagonal element, column-wise (regret) or row-wise (generalization error).

Note that these numbers represent a general observation that the regret is asymmetric, see for instance Fig. 4 and prior works [Gib+20; KSF14; GCB18; BBB18; Gib+20].

Regret informs us about the severity of the CSM problem since it assesses that when inspect a given dataset, how much detection performance can be impacted depending on the dataset used for training a given detector.
Interestingly, when the sources differ in a single processing step, the regret allows evaluating the impact of this step on the CSM problem.

The regret reports about the test source, and is computed using two detectors. Generalization error is an alternative metric, computed with one detector trained on one single source but tested over two sources.

Definition 8 [Generalization error] The generalization error $R_{XY,X}$ is the difference between the inconsistency ϵ_{XY} and the intrinsic difficulty ϵ_X, as shown in Eq. 2,

$$R_{XY,X} = \epsilon_{XY} - \epsilon_X. \tag{2}$$

Both the regret and the generalization error are still dependent on the magnitude of intrinsic difficulty. But their nature and their uses are very different. Roughly speaking, generalization error focus more on the capacity of a detector used over different sources. On the opposite, regret reports how much a given testing source is sensitive to the choice of the training source.

Eventually, it is often interesting to normalize a metric; in the present case of the CSM problem, the *regret* can

be normalized in order to contrast with the intrinsic difficulty of the source.
Note that this also allows including all prior works results in this survey to propose an exhaustive terminology. To this end we introduce the *relative regret*:

Definition 9 [Relative regret] The relative regret is the regret normalized by the intrinsic difficulty, shown in Eq.3:

$$\frac{\epsilon_{XY} - \epsilon_Y}{\epsilon_Y}. \tag{3}$$

Our definitions, based on [GCB18; ŠAP22], aim to solve the discrepancy of the terms "*inconsistency*" and "*regret*" [GCB18; Gib+20; Abe+22; Abe+23].

Limitations Measuring the CSM problem via detector error-rate stems from an operational approach: it focuses on the symptoms of CSM, but heavily depends on a type of detector used, which makes regret-based mitigation difficult to generalize to different detectors.

Reporting the regret also depends on the chosen performance metric. The most common choice in steganalysis is the probability of error P_E under equal prior assumption. However some alternatives have also been proposed such as the miss-detection rate for a prescribed false-alarm and the Area Under ROC Curve (AUC) to cite few.

4.2 Measuring a mismatch via distance in a feature space

The second option to quantify the CSM problem is as a distance of cover sources projected to a lower-dimensional feature space. This relates the definition of Cachin's steganographic security [Cac98], defined as a distance between hypothetical distributions between covers and stegos.

Definition 10 [Feature space] Feature space is a collection of n variables, extracted using the same function $\phi(\cdot)$, from different objects and carefully designed for a specific goal of analysis. Indeed, features aim at preserving statistical properties of objects while normalizing their representation, reducing the dimensionality and hence helping the desired analysis.

The feature spaces can be defined manually, for instance handcrafted steganalysis features DCTR [HF14] or GFR [Son+15], or constructed ad-hoc by a trainable component.

The distance can be measured with Euclidean norm ℓ^2, or Kullback-Leibler divergence (KLD); a popular metric for domain adaptation (Sec. 7.4) is the maximum-mean discrepancy (MMD).

Definition 11 [MMD] Maximum-mean discrepancy (MMD) is an average-link distance metric between cover sources X and Y, defined in Eq. 4,

$$\begin{aligned}\mathrm{MMD}^2(X,Y) &= \langle \mu_X - \mu_Y, \mu_X - \mu_Y \rangle = \\ &= \langle \mu_X, \mu_X \rangle + \langle \mu_Y, \mu_Y \rangle - 2\langle \mu_X, \mu_Y \rangle,\end{aligned} \tag{4}$$

where $\mu_X = \mathbb{E}_X[\phi(x)]$, and $\langle \mu_X, \mu_Y \rangle = \mathbb{E}_{X,Y}[k(x,y)]$, given a feature extractor $\phi(\cdot)$ and a kernel $k(\cdot)$.

Recently introduced metrics between cover sources are chordal distance [Abe+23] and KLD based on the covariance matrix with regret [Mal+23].

Limitations Ideally, the feature distance should correlate to the regret [Abe+23; Mal+23]. However while distances are symmetric $d(X,Y) = d(Y,X)$ this generally does not hold true for the regret in steganalysis which is asymmetric, as shown in Fig. 4. For MMD and ℓ^2, both symmetric metrics, this is clearly incorrect.
Recent works addressed the CSM problem by relating distances in features-space with the detection regrets with the goal to understand better the causes of the CSM problem. These approaches, however, are still in their infancy and faces fundamental challenges.

5 Bibliometry

This section introduces the methods used to answer the research questions from Sec. 1. Sec. 5.1 presents the collection of literature (RQ1). Sec. 5.2 describes measuring the research trend (RQ1). Sec. 5.3 depicts how the literature results were aggregated (RQ2, RQ3).

From now on, the survey focuses on the specific case of image steganalysis. Indeed, the state of the art in other types of covers, such as video, audio, text files, etc. are much less developed. In particular, the CSM problem, even though it is studied by some papers, for instance in video steganalysis [LSZ23] or audio steganalysis [Lin+20], is not enough covered.

5.1 Collection of literature

The collection strategy consists of (1) initial sampling, (2) identification of relevant papers, and (3) breadth-first search using forward and backward references.

Sampling The initial samples were acquired with a query "cover source mismatch steganalysis" from two metasearch engines, namely Google Scholar (GS) and DBLP, and independently from the editor or publisher. We take the first 50 results, sorted by relevance with no additional filters.

Identification Each paper was proofread to identify whether it is relevant. Relevant papers fit into at least one of the following criteria:

- Paper discusses the effects of CSM.
- Paper reports results with CSM.
- Paper investigates the impact of factors on CSM.
- Paper attempts to mitigate CSM.

Only a few papers mentioning the CSM were in fact excluded, for instance, when CSM is a future work.

Reference search For each relevant paper, we searched its references (*forward search*) as well as papers citing it using the GS feature "Cited by" (*backward search*). On the newly found papers, we applied the same reference search, proceeding in the breath-first order until exhaustion.

In the end, we collected and annotated 102 papers. The complete annotated bibliography is presented in Appendix A.

5.2 Measuring the Research Trend

Each paper in the bibliography is labelled with tags, denoting assumed or explored topics. The tags allow for measuring the trends of CSM research, separately for causes and mitigation. The number of citations is acquired from GS, as of 5 October 2023. For each paper, we compute the *topic coverage* (Eq. 5),

$$\frac{1}{\#\text{tags}}. \tag{5}$$

5.3 Aggregation of results

Although a comparison of the results in the literature is hardly possible due to different experimental setups, a relative measure may account for it to some extent. We convert the results to relative regret (Eq. 3), and aggregate by taking the expectation across cover sources, according to Eq. 6,

$$\mathbb{E}_{X,Y}\left[\frac{\epsilon_{XY}-\epsilon_Y}{\epsilon_Y}\right]. \tag{6}$$

Relative regret marginalizes differences of experimental setups, and can be extracted from the existing literature, when inconsistency or regret are reported together with the source's intrinsic difficulty.

However, the statistics should match in error metrics; problematic is also dynamic threshold used in P_E. Furthermore, relative regret assumes that CSM affects performance proportionally.

6 Causes of CSM

In this section, we answer RQ 2 for image covers.

6.1 Image cover source

A cover is defined by the whole imaging pipeline, which consists in the acquisition (from the scene up to the electrical signal) and processing (from electrical signals to the image file) [Ram+05; HB23]. The diversity of the cover sources stems from the variety in acquisition devices and parameters, processing software and their specific operations and parameters. But it is also enriched by the initial captured content and the later compressions applied by specific software, such as social networks.

This diversity makes it hard to give a proper all-encompassing model of the cover source. However it is possible to measure the impact of each operation. To get their impact, we gather the intrinsic difficulties and the source inconsistencies from [BBB18; YKF18; Lin+18; Gib+20; BHB22]. We use them to compute the relative source regrets using Eq. 3. We then compute the expected relative regrets using Eq. 6 gathering every causes into 6 categories: content, device, colour, filter, resize and JPEG.

The distribution of the relative regrets for each step is reported in Fig. 5. As expected, the JPEG compression step has the biggest impact on CSM. The impact of processing steps tends to be higher the later they are in the pipeline.

Sec. 6.2 summarizes the state of the art using this 6-step categorization of the cover source steps.

6.2 Coverage of Causes

Each and every step discussed below is an arbitrary division of all the gathered causes of CSM. Readers should be aware, for instance, that filtering operations can be performed at different steps of the pipeline. Equally, colour operations can be performed at a later stage. Finally, while we generally assume that the JPEG compression comes last, it does not guarantee that later processing won't induce CSM.

Despite these nuances, we chose, for the sake of clarity in answering RQs. 1-2, to present a summary in the general order in which each cause appears in the overall pipeline. We strongly recommend reading the studies to get the full explanation of the reported findings.

Content While not properly a part of the processing pipeline, the content has long been recognized a cause of heterogeneity in steganalysis [KSM06]. The level of texture, also called *texture complexity* (TC), allows

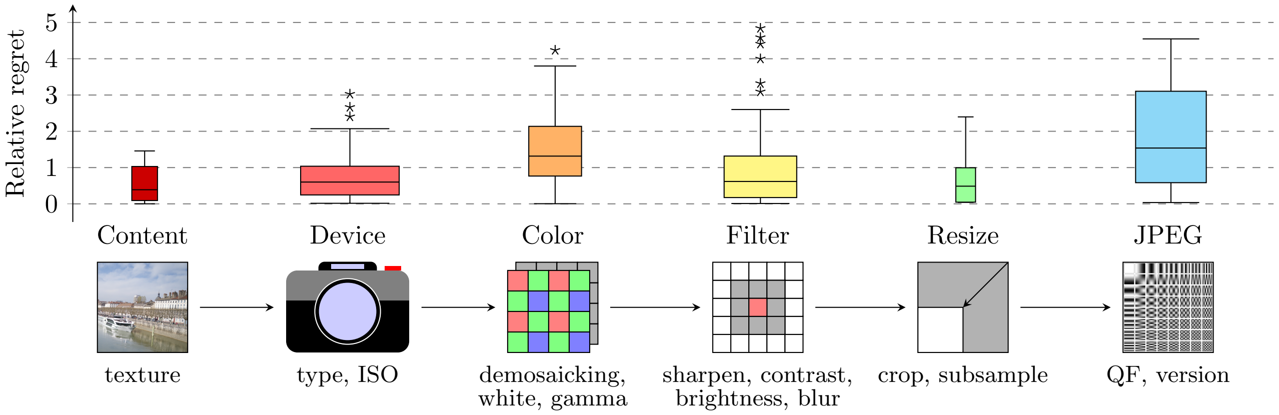


Figure 5: Generic schema of image processing pipeline (IPP), together with the impact of each step on CSM.

quantification of image similarity. Such a metric was proposed in [Hu+17] Datasets with high TC are harder to train on [Gib+20; Yu+23], but generalize better on testing sets with low TC than the other way around [Hu+17; Hu+19].

The amount of textures depends on the captured scene, but later processing, such as filtering or resampling, greatly impact the final textures as well.

Device The device model is a common source of diversity in the literature [BFP11; CGB19]. Research also covers the impact of individual device [Lin+18], and acquisition parameters: ISO sensitivity (moderate impact), aperture, and exposure time (low impact) [Gib+20].

Colour Colour processing involves demosaicking, i.e., removing the Bayer grid from the raw image by interpolating colours, and colour balancing, rendering white and gray shades. Optional steps are linear colour correction, and the non-linear gamma correction. Diversity in demosaicking algorithm is common in the literature [CGB19], even though its impact is lower than filtering or resampling [Gib+20; Abe+22]. On the other hand, relative excess regrets extracted from [BBB18] are high. This study reports noticeably low ϵ_{XY}, which explains the high relative impact of colour step in Fig. 5. For all these processing in particular, Fig. 5 would benefit in having more results.

Filter Filtering encompasses neighbourhood-based processing, such as denoising [NO08], unsharpening [PRM00], blurring, sharpening, or edge enhancement. These operations are shown to have a strong impact on the CSM [BF17; Gib+20; Gib+22; Abe+22]. Due to the large number of filtering parameters, measuring the CSM quickly becomes a computationally overwhelming task.

Resize Resizing can be carried out with two very distinct operations: cropping and resampling. Cropping consists merely in removing pixels row or column without any additional modification. While this may induce a shift in the grid, such as Bayer or JPEG, it completely preserves statistics in pixels domain. On the opposite, resampling requires a low-pass digital filtering operation (to interpolate missing values) hence reduces the texture complexity and is a strong CSM factor. Downsampling can reduce CSM between mismatching cover sources [Zha+19]. Both cropping and resampling were used in ALASKA [CGB19].

JPEG JPEG compression is indisputably the most impactful cause of CSM. The constant attention from the community [GFH06; KSF14; Kon+16; CGB20] focused mostly on the quality factor (QF) controlling the distortion-rate tradeoff. CSM is also induced by double compression [Rod+22] or JPEG implementation [BHB22]

Fig. 6 answers the RQ. 2. It shows the evolution of the number of papers, from 2006 onward, studying the causes of CSM. To the 6 categories that we consider in the paper, we added 3 others, often found in studies. First, “IPP” designates studies with unspecified cover source processes, but still use them as a cause of CSM. Then “Stego” refers to papers dealing with SSM, and “Dataset” to CSM between datasets. Additionally, when a paper dealt with more than one cause, we chose to split its weight evenly.

Note that the number of papers used in Fig. 6 is less than the total number used in the survey. It is because not all papers accurately discussed the causes of CSM.

Heuristics for CSM Impacts To conclude this section, let us mention some complementary papers that use heuristics to approximate the impact on CSM.

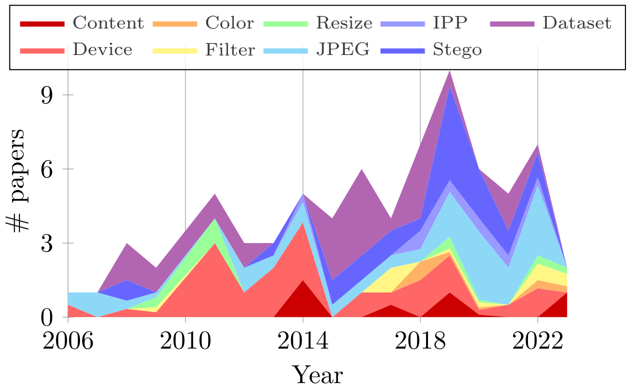


Figure 6: Area chart capturing timeline of CSM causes.

A heuristic close enough to the detector error can facilitate mitigation of CSM caused by the parameter whose impact is being approximated. For instance, [Hu+17] designs a similarity for texture complexity in images. [YF20] suggests a metric for JPEG quantization tables (QT)[6].

7 Mitigation of CSM

This section answers RQ3. We start by the idealistic case – clairvoyant scenario (Sec. 7.1), followed by major mitigation schools: atomistic (Sec. 7.2), holistic (Sec. 7.3), and domain adaptation (Sec. 7.4). Sec. 7.5 summarizes miscellaneous other techniques. In Sec. 7.6 we analyse the trends in the research.

7.1 Clairvoyant scenario

Perfect mitigation of CSM is possible, if steganalyst gets to know the cover source, and trains the detector on it. Such simple strategy, referred to as the clairvoyant scenario [Pev11], comes from a conservative interpretation of Kerckhoff's principle, according to which what is not secret is assumed to be public.

Limitations However, knowing the cover source is applicable only sometimes in practice. It assumes the steganographer publishes the cover source, either deliberately, unintentionally, or forced by an active steganalyst. Otherwise, CSM is present, and steganalyst must seek different ways to mitigate it. The current state of the art has three major schools: atomistic, holistic and domain-adaptation.

7.2 Atomistic steganalysis

A natural extension of the clairvoyant scenario to unknown cover sources is to train multiple detectors on different cover sources, and choose the correct detector for the input image. Such *atomistic* detector, shown in Fig. 8a, involves two steps:

1. the forensic step (*select*), which identifies the cover source from the input image; and
2. the steganalysis step (*detect*), a pool of detectors trained on different cover sources.

The training consists of selecting the cover sources and training the selector and one detector per cover source. During the testing, the input cover source is determined, and the input is fed into the associated detector. If the detector for the input cover source is available, the atomistic detector performs as well as in the clairvoyant scenario [Hou+12; Zen+15].

Selector A major question is how to construct the newly introduced selector component. Cover source identification typically uses a feature space and an unsupervised [Zen+15; Hou+14; PBC14] or non-parametric [Gom+18] construction. The cover source can also be partially reconstructed using the means of forensic analysis [BCE10; Hou+12].

Limitations The first limitation is that the number of possible cover sources is intractable. The selector must be able to deal with the situation when the cover source is not present in the pool.

The second problem is that the success of the atomistic detector relies on the selector. Errors of the selector add up with the errors of the detectors [ŠAP22].

The third issue is the selection the cover sources to train on. The aim of the steganalyst is to a good coverage of the source domain, but the more detectors one wants, the longer the training time becomes [Abe+23].

7.3 Holistic steganalysis

A detector performs the best on the cover source seen during the training. When trained on multiple cover sources, the performance is usually slightly worse than of the dedicated per-source detectors. By contrast, the performance degrades far more on unseen cover sources. The amount of degradation depends on how different each cover source is from the training set.

A *holistic* approach gives up on matching the clairvoyant performance and training a large number of dedicated detectors. Instead, one detector is trained on a heterogenous dataset with a large amount of

[6] QT metric weights high frequencies unintuitively. E.g., standard QF75 is closer to QF76 than to QF75 with incremented DC, although 90% values differ.

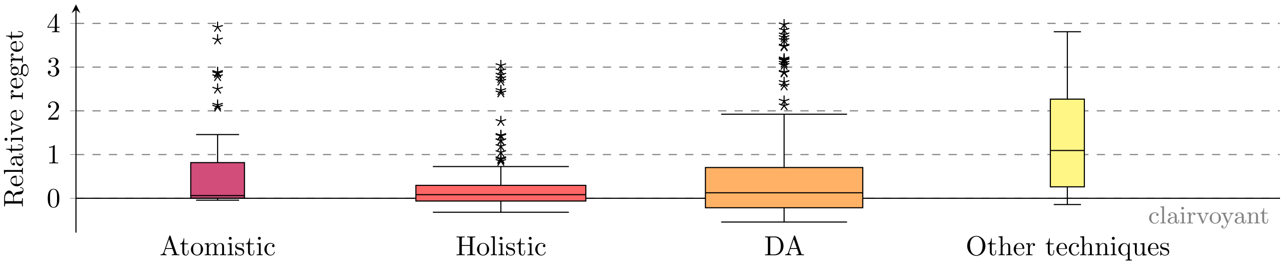
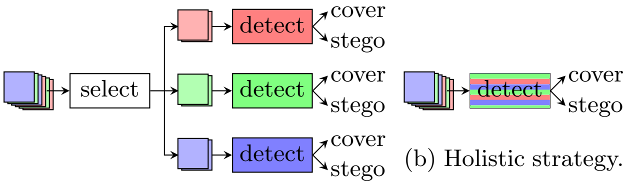


Figure 7: Effect of existing mitigation strategies on CSM, compared to clairvoyant scenario.



(a) Atomistic strategy.

(b) Holistic strategy.

Figure 8: Strategies to mitigate cover-source mismatch.

cover sources, as shown in Fig. 8b. The idea is that a sufficient coverage over sources ensures a good performance of the detector on unseen sources, and leads to mitigation of CSM.

Heterogeneity The major challenge in holistic approach is the construction of the training database. A number of sources certainly relate to the robustness of the dataset. Research suggests that training on fewer, carefully chosen cover sources yields better results, than on a huge number of blindly collected cover sources [Xu+15; Abe+22].

Model selection Architecture of holistic detector is critical for model performance. Not only is the detector expected to detect steganography, but it also needs to do so in heterogeneous environment [Fri+11]. Increasing flexibility of a model can lead to greater overfitting, which can be mitigated by increasing the dataset size, or by regularization techniques [MK13; Ng+14].

Limitations The first limitation is that the holistic training requires a very large dataset [LK12; PBC14]. The second problem is that the holistic approach performs worse than atomistic approach on a fixed set of cover sources [Hou+12; Zen+15]. The third issue is that the performance is very sensitive to the cover sources covered in the training set. The success of the mitigation also strongly depends on the selection of the detector architecture.

7.4 Domain adaptation

The mitigation of atomistic and holistic approaches is limited by a number of sources used during training. The idea of domain adaptation (DA) is training on a single cover source called *source domain*, and use the learnt knowledge to adapt to an unseen cover source called *target domain*. This is done by aligning the domains, illustrated in Fig. 9b by coloured circles, so that the detector can better generalize to a diverse domain.

The general DA workflow illustrated in Fig. 9a is. (1) converting the test sample to the adapted feature space; and (2) passing them to a detector, trained in this feature space.

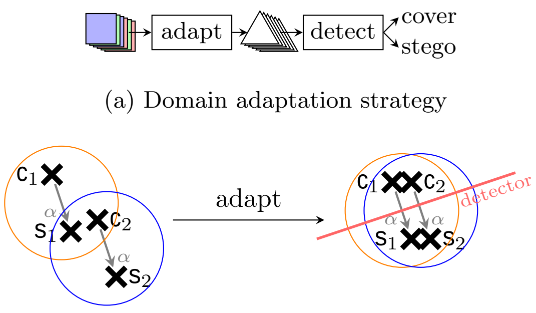
Adaptation The feature space should align different cover sources close, yet maximize the shift caused by steganography. Its construction is challenging, because the labels for the target domain are usually not available. Literature presents solutions using various unsupervised techniques, such as clustering or manifold alignment [Li+13; LM16a; Kon+16; Fen+17; Jia+20; Abe+21], possibly aided by guiding features and pseudo-label prediction [Zha+21; Zha+22].

Limitations Existing methods only extract specific statistics, such as means or higher moments, which may be insufficient to mitigate CSM over different cover sources. Moreover, using handcrafted steganalysis features for domain adaptation may lead to unsuccessful mitigation, because although sensitive to steganography, they are also affected by CSM [Yan+21].

7.5 Other techniques

Apart from the three major approaches, we identify other, marginally explored, techniques.

Re-embedding Re-embedding into the test images increases the spread of the stego shift directly in the target domain. Training multiple detectors,



(a) Domain adaptation strategy

(b) Domain adaptation schema

Figure 9: Domain adaptation strategy.

similar to diffusion models, i.e., on cover-stego, stego-double stego, etc., may further improve the detection performance [LM16b; LM18; LM19; Yu+23].

Feature projection The second limitation of DA mentioned in Sec. 7.4 may be tackled by modifying the features, so that the projected features are insensitive to cover source heterogeneity [PK13; Xue+19]. Feature engineering w.r.t. CSM is not trivial, for instance, fusing feature sets, which helps in the matched scenario, may degrade performance in the presence of CSM [Fri+11].

One-class detector Figs. 1, 3 and 9b depict linear detectors, but there is a variety of detectors, which in some cases may perform better or be more robust to CSM, such as non-linear classifiers or one-class classifiers [Pev08].

Unsupervised Learning A steganography may be treated as an outlier, detected using unsupervised learning. This is particularly common in pooled steganalysis, where steganalyst looks for guilty steganographer among a pool of communicating actors. The techniques used are outlier detection, clustering based on MMD [KP11; KP12b] or k-nearest neighbours [HZX16].

7.6 Trends in Research on CSM Mitigation

In a similar fashion to the CSM causes in Fig. 6, we show the trend of the mitigation strategies in Fig. 10, in terms of the number of papers per year. The number of papers has been growing until peak in 2018-2020, and descending since.

Holistic and atomistic strategies appear consistently over the entire time period. Holistic strategy is the most common, which can be explained by better generalization capabilities, as demonstrated by being the winning strategy both in BOSS and ALASKA competitions. Unsupervised approach appears mainly in the pooled steganalysis literature. The domain adaptation appears more in the recent years over which we have seen the dominance of domain adaptation, connected with the popularity of deep learning.

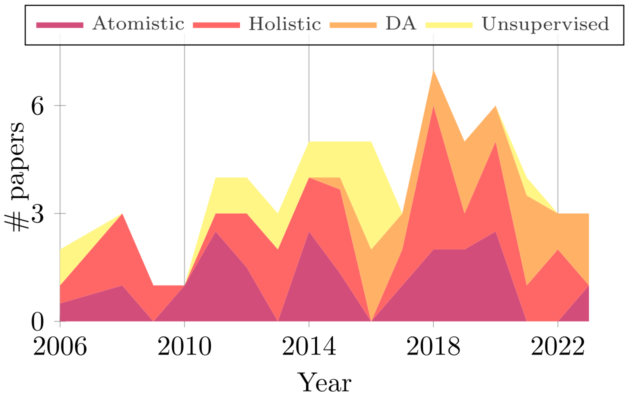


Figure 10: Area chart capturing timeline of CSM mitigation strategies, with y-axis step 3.

8 Discussion

Steganalysis competitions BOSS and ALASKA had a profound impact on the field, and stimulated a lot of new energy. Their experimental setups were followed by the research long after they ended. Future competitions should be carefully designed to facilitate comparisons, such as the one carried out in this survey; of importance to the CSM are the factors of diversity, steganographic schemes, and performance metrics.

Open Question 1: Have all the causes of CSM and SSM been clearly identified ? Are their effects well measured ?

Pushing forward in this direction, the steps of the IPP are usually studied in isolation, but their order might impact the CSM. Interactions between the steps exist, e.g., between resampling and sharpening [Abe+22] or between SSM and the cover source [Rei+19].

Open Question 2: How can we measure the crossed effect of different causes on the CSM ?

Answering this question, among others, challenges the community to build methods that will need to deal with the computational complexity of the task.

We detailed the different approaches to measure CSM, highlighting the limitations in each one, but the question is still mostly an open one. As the current tools are symmetric, a "theoretical" measure of CSM can only poorly relate to the practical regret. Designing better tools, in order to avoid the bias of training a detector, is probably one of the most promising steps at characterizing the CSM in the short term.

Open Question 3: How is CSM related with statistical properties of the cover sources ?

Finally, as already mentioned, this survey is focused on the case of steganalysis natural images. One question that we can naturally draw from our present research would be to see if the same observations can be made for other types of digital media.

Open Question 4: How does CSM impact steganalysis of other types of covers (audio, video, . . .) ?

Existing methods to measure the CSM impact depend on the detector, or on the feature space. Suitability of these assumptions, as well as possible alternatives, is to be investigated. Meta-research, such as result aggregation in Fig. 5, would benefit if future studies provided intrinsic difficulties and inconsistencies.

Existing mitigation strategies may aid at solving the problem, yet fail in pessimistic scenarios, such as unknown processing history. The atomistic approach is suitable for a closed set of cover sources, but fails on open-set problems where holistic performs better. The increased popularity of domain adaptation correlates with the introduction of deep learning in steganalysis.

We strive to sample and aggregate the existing literature on CSM objectively. However, we are aware of potential biases: (1) sampling bias due to search engine ranking, isolation of papers from the rest of literature, or incorrect assessment of relevance, (2) bias due to incorrect annotation of the paper, and (3) bias of impact estimates, when the paper results cannot be used for aggregation, e.g., when reported via graph.

Adversarial scenario for CSM CSM is usually understood as a problem for steganalyst. It can also be interpreted as a game between the steganographer and the steganalyst, a modified rock-paper-scissors, where the steganalyst wins on match and steganographer on the mismatch of the shapes [Fri09; GPK23]. This approach has been tested in a very constrained scenario so far.

9 Conclusion

The research in the last 20 years has been looking into the cover-source mismatch problem from various directions. Many strategies to suppress CSM exist, but the core of the problem is still present. Successful mitigation must go hand in hand with understanding of the causes. Reliable mitigation of CSM and SSM is essential for operational universal steganalysis.

10 Abbreviations

The following abbreviations are used in this manuscript, given in alphabetical order.

- AUC: area under curve
- BOSS: break our steganographic system.
- CORAL: correlation alignment
- CSM: cover-source mismatch.
- DA: Domain Adaptation
- DBLP: digital bibliography and library project
- DCT: discrete cosine transform
- DCTR: DCT residual
- GFR: Gabor filter residual
- GS: Google Scholar
- IPP: image processing pipeline.
- KLD: Kullback-Leibler divergence
- LSB: least-significant bit
- MCM: mechanical component monitoring
- MI: Medical Imaging.
- ML: machine learning.
- MMD: maximum-mean discrepancy
- NLP: natural language processing
- QF: quality factor
- ROC: reciever-operating characteristic
- RQ: research question.
- SR: speech recognition
- SSM: stego-scheme mismatch
- TC: texture complexity
- TCA: transfer component analysis
- TR: temporal reasoning

References

[Abe+21] Rony Abecidan et al. “Unsupervised JPEG Domain Adaptation for Practical Digital Image Forensics”. In: *WIFS*. IEEE. 2021, pp. 1–6.

[Abe+22] Rony Abecidan et al. “Using Set Covering to Generate Databases for Holistic Steganalysis”. In: *WIFS*. IEEE. 2022, pp. 1–6.

[Abe+23] Rony Abecidan et al. “Leveraging Data Geometry to Mitigate CSM in Steganalysis”. In: *WIFS*. IEEE. 2023, pp. 1–5.

[BBB18] Dirk Borghys, Patrick Bas, and Helena Bruyninckx. “Facing the Cover-Source Mismatch on JPHide using Training-Set Design”. In: *IH&MMSec*. ACM, 2018, pp. 17–22.

[BCE10] Mauro Barni, Giacomo Cancelli, and Annalisa Esposito. “Forensics-Aided Steganalysis of Heterogeneous Images”. In: *ICASSP*. IEEE. 2010, pp. 1690–1693.

[BDP07] John Blitzer, Mark Dredze, and Fernando Pereira. “Biographies, Bollywood, Boomboxes and Blenders: Domain Adaptation for Sentiment Classification”. In: *ACL*. 2007, pp. 440–447.

[BF17] Mehdi Boroumand and Jessica Fridrich. “Scalable Processing History Detector for JPEG Images”. In: *EI* 29 (2017), pp. 128–137.

[BFP11] Patrick Bas, Tomáš Filler, and Tomáš Pevný. “”Break Our Steganographic System”: the Ins and Outs of Organizing BOSS”. In: *IH*. Springer. 2011, pp. 59–70.

[BHB22] Martin Beneš, Nora Hofer, and Rainer Böhme. “The Effect of the JPEG Implementation on the Cover-Source Mismatch Error in Image Steganalysis”. In: *EUSIPCO*. IEEE. EURASIP, 2022, pp. 1057–1061.

[Cac98] Christian Cachin. “An Information-theoretic Model for Steganography”. In: *IH*. Springer. 1998, pp. 306–318.

[Can+08] Giacomo Cancelli et al. “A Comparative Study of ±1 Steganalyzers”. In: *MMSP*. IEEE. 2008, pp. 791–796.

[CF15] Rémi Cogranne and Jessica Fridrich. “Modeling and Extending the Ensemble Classifier for Steganalysis of Digital Images Using Hypothesis Testing Theory”. In: *TIFS* 10.12 (2015), pp. 2627–2642.

[CGB19] Rémi Cogranne, Quentin Giboulot, and Patrick Bas. “The ALASKA Steganalysis Challenge: A First Step Towards Steganalysis”. In: *IH&MMSec*. ACM, 2019, pp. 125–137.

[CGB20] Rémi Cogranne, Quentin Giboulot, and Patrick Bas. “ALASKA#2: Challenging Academic Research on Steganalysis with Realistic Images”. In: *WIFS*. IEEE. 2020, pp. 1–5.

[Che+17] Mo Chen et al. “JPEG-Phase-Aware Convolutional Neural Network for Steganalysis of JPEG Images”. In: *IH&MMSec*. ACM, 2017, pp. 75–84.

[Cog15] Rémi Cogranne. “A Sequential Method for Online Steganalysis”. In: *WIFS*. IEEE. 2015, pp. 1–6.

[CSF17] Rémi Cogranne, Vahid Sedighi, and Jessica Fridrich. “Practical strategies for content-adaptive batch steganography and pooled steganalysis”. In: *ICASSP*. IEEE. 2017, pp. 2122–2126.

[DBF16] Tomáš Denemark Denemark, Mehdi Boroumand, and Jessica Fridrich. “Steganalysis features for content-adaptive JPEG steganography”. In: *TIFS* 11.8 (2016), pp. 1736–1746.

[Fen+17] Chaoyu Feng et al. “Contribution-Based Feature Transfer for JPEG Mismatched Steganalysis”. In: *ICIP*. IEEE. 2017, pp. 500–504.

[Fri+11] Jessica Fridrich et al. “Breaking HUGO–the Process Discovery”. In: *IH*. Springer. 2011, pp. 85–101.

[Fri09] Jessica Fridrich. *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, 2009. Chap. 10.

[GCB18] Quentin Giboulot, Rémi Cogranne, and Patrick Bas. “Steganalysis into the Wild: How to Define a Source?” In: *MWSF*. Vol. 30. 7. SPIE. IS&T, 2018, pp. 1–12.

[GFH06] Miroslav Goljan, Jessica Fridrich, and Taras Holotyak. “New Blind Steganalysis and Its Implications”. In: *SSWMC*. Vol. 6072. SPIE. 2006, p. 607201.

[Gib+20] Quentin Giboulot et al. “Effects and Solutions of Cover-Source Mismatch in Image Steganalysis”. In: *SPIC* 86 (2020), p. 115888.

[Gib+22] Quentin Giboulot et al. “The Cover Source Mismatch Problem in Deep-Learning Steganalysis”. In: *EUSIPCO*. IEEE. EURASIP, 2022, pp. 1032–1036.

[GL21] Hao Guan and Mingxia Liu. “Domain Adaptation for Medical Image Analysis: a Survey”. In: *TBME* 69.3 (2021), pp. 1173–1185.

[GL22] Hao Guan and Mingxia Liu. “Domain Adaptation for Medical Image Analysis: A Survey”. In: *TBME* 69.3 (2022), pp. 1173–1185.

[Gom+18] François Kasséné Gomis et al. “Multiple Linear Regression for Universal

Steganalysis of Images". In: *ISCV*. IEEE. 2018, pp. 1–4.

[Gon95] Yifan Gong. "Speech Recognition in Noisy Environments: A Survey". In: *Speech Communication* 16.3 (1995), pp. 261–291.

[GPK23] Quentin Giboulot, Tomáš Pevný, and Andrew Ker. "The Non-Zero-Sum Game of Steganography in Heterogeneous Environments". In: *TIFS* 18 (2023), pp. 4436–4448.

[HB23] Keigo Hirakawa and Farhan Baqai. *Digital Camera Processing Pipeline*. The Wiley-IS&T Series in Imaging Science and Technology. Wiley, 2023. ISBN: 9780470686096.

[HF14] Vojtěch Holub and Jessica Fridrich. "Low-complexity Features for JPEG Steganalysis Using Undecimated DCT". In: *TIFS* 10.2 (2014), pp. 219–228.

[Hou+12] Xiaodan Hou et al. "Forensics-aided Steganalysis of Heterogeneous Bitmap Images with Different Compression History". In: *MINES*. IEEE. 2012, pp. 874–877.

[Hou+14] Xiaodan Hou et al. "A Novel Steganalysis Framework of Heterogeneous Images Based on GMM Clustering". In: *SPIC* 29.3 (2014), pp. 385–399.

[Hu+17] Donghui Hu et al. "A Study of the Two-way Effects of Cover-Source Mismatch and Texture Complexity in Steganalysis". In: *IWDW*. Springer. 2017, pp. 601–615.

[Hu+19] Donghui Hu et al. "Study on the Interaction between the Cover-Source Mismatch and Texture Complexity in Steganalysis". In: *MTAP* 78 (2019), pp. 7643–7666.

[Hus+20] Israr Hussain et al. "A Survey on Deep Convolutional Neural Networks for Image Steganography and Steganalysis". In: *TIIS* 14.3 (2020), pp. 1228–1248.

[HZX16] Xiaodan Hou, Tao Zhang, and Chen Xu. "New Framework for Unsupervised Universal Steganalysis via SRISP-Aided Outlier Detection". In: *SPIC* 47 (2016), pp. 72–85.

[Jia+20] Ju Jia et al. "Transferable Heterogeneous Feature Subspace Learning for JPEG Mismatched Steganalysis". In: *Pattern Recognition* 100 (2020), p. 107105.

[KDF23] Edgar Kaziakhmedov, Eli Dworetzky, and Jessica Fridrich. "Observing Bag Gain in JPEG Batch Steganography". In: *WIFS*. IEEE, 2023.

[Ker+13] Andrew Ker et al. "Moving Steganography and Steganalysis from the Laboratory into the Real World". In: *IH&MMSec*. ACM, 2013, pp. 45–58.

[Ker06] Andrew Ker. "Batch Steganography and Pooled Steganalysis". In: *IH*. Springer. 2006, pp. 265–281.

[Kon+16] Xiangwei Kong et al. "Iterative Multi-Order Feature Alignment for JPEG Mismatched Steganalysis". In: *Neurocomputing* 214 (2016), pp. 458–470.

[KP11] Andrew Ker and Tomáš Pevný. "A New Paradigm for Steganalysis Via Clustering". In: *MWSF*. Vol. 7880. SPIE. 2011, pp. 312–324.

[KP12a] Andrew Ker and Tomáš Pevný. "Batch Steganography in the Real World". In: *MMSec*. ACM, 2012, pp. 1–10.

[KP12b] Andrew Ker and Tomáš Pevný. "Identifying a Steganographer in Realistic and Heterogeneous Data Sets". In: *MWSF*. Vol. 8303. SPIE. IS&T, 2012, pp. 182–194.

[KP14a] Andrew Ker and Tomáš Pevný. "A Mishmash of Methods for Mitigating the Model Mismatch Mess". In: *MWSF*. Vol. 9028. SPIE. IS&T, 2014, pp. 189–203.

[KP14b] Andrew Ker and Tomáš Pevný. "The Steganographer is the Outlier: Realistic Large-Scale Steganalysis". In: *TIFS* 9.9 (2014), pp. 1424–1435.

[KP16] Stefan Katzenbeisser and Fabien Petitcolas. *Information Hiding*. Artech house, 2016. Chap. 3.

[KSF14] Jan Kodovský, Vahid Sedighi, and Jessica Fridrich. "Study of Cover Source Mismatch in Steganalysis and Ways to Mitigate its Impact". In: *MWSF*. Vol. 9028. SPIE. IS&T, 2014, pp. 204–215.

[KSM05] Mehdi Kharrazi, Husrev Sencar, and Nasir Memon. "Benchmarking Steganographic and Steganalysis Techniques". In: *SSWMC*. Vol. 5681. SPIE. 2005, pp. 252–263.

[KSM06] Mehdi Kharrazi, Husrev Sencar, and Nasir Memon. "Performance Study of Common Image Steganography and Steganalysis Techniques". In: *EI* 15.4 (2006), pp. 041104–041104.

[Lea+22] Vaila Leask et al. "UNCOVER: Development of an Efficient Steganalysis Framework for Uncovering Hidden Data in Digital Media". In: *ARES*. ACM, 2022.

[Li+13] Xiaofeng Li et al. "Generalized Transfer Component Analysis for Mismatched JPEG Steganalysis". In: *ICIP*. IEEE. 2013, pp. 4432–4436.

[Lin+18] Li Lin et al. "Domain Adaptation in Steganalysis for the Spatial Domain". In: *MWSF* 2018.7 (2018), pp. 319–1.

[Lin+20] Yuzhen Lin et al. "Tackling the Cover-source Mismatch Problem in Audio Steganalysis with Unsupervised Domain Adaptation". In: *SPL* 28 (2020), pp. 1475–1479.

[LK11] Ivans Lubenko and Andrew Ker. "Steganalysis Using Logistic Regression". In: *MWSF*. Vol. 7880. SPIE. 2011, pp. 193–203.

[LK12] Ivans Lubenko and Andrew Ker. "Going from Small to Large Data in Steganalysis". In: *MWSF*. Vol. 8303. SPIE. IS&T, 2012, pp. 172–181.

[LK19] Chang Liu and Matthias Kirchner. "CNN-based Rescaling Factor Estimation". In: *IH&MMSec*. ACM, 2019, pp. 119–124.

[LM16a] Daniel Lerch-Hostalot and David Megías. "Manifold-Alignment Approach to Cover-Source Mismatch in Steganalysis". In: *RESCI* (2016).

[LM16b] Daniel Lerch-Hostalot and David Megías. "Unsupervised Steganalysis based on Artificial Training Sets". In: *Engineering Applications of Artificial Intelligence* 50 (2016), pp. 45–59.

[LM18] Daniel Lerch Hostalot and David Megías Jiménez. "Diagnóstico de CSM en Estegoanálisis". In: *RECSI* (2018).

[LM19] Daniel Lerch-Hostalot and David Megías. "Detection of Classifier Inconsistencies in Image Steganalysis". In: *IH&MMSec*. 2019, pp. 222–229.

[LM23] Daniel Lerch-Hostalot and David Megias. "Real-World Actor-Based Image Steganalysis via Classifier Inconsistency Detection". In: *ARES*. ACM, 2023.

[LSZ23] Verena Lachner, Katharina Schaar, and Ralf Zimmermann. "CSM in Motion Vector Steganalysis: The Effect of Coders on Motion Vectors in H.264 Video Encoding". In: *ICASSP*. IEEE. 2023, pp. 1–5.

[Luo+17] Yuan Luo et al. "Natural Language Processing for EHR-based Pharmacovigilance: a Structured Review". In: *Drug Safety* 40 (2017), pp. 1075–1089.

[Mal+23] Antoine Mallet et al. "Identification de Développements d'Images par Matrices de Corrélations". In: *XXIXème Colloque Francophone de Traitement du Signal et des Images*. GRETSI'23. Université de Grenoble and Association Gretsi. 2023.

[MK13] Julie Makelberge and Andrew Ker. "Exploring Multitask Learning for Steganalysis". In: *MWSF*. Vol. 8665. SPIE. 2013, pp. 218–227.

[Ng+14] Wing Ng et al. "Steganalysis Classifier Training Via Minimizing Sensitivity for Different Imaging Sources". In: *Information Sciences* 281 (2014), pp. 211–224.

[NO08] Truong Nguyen and Soontorn Oraintara. "The Shiftable Complex Directional Pyramid—Part II: Implementation and Applications". In: *TSP* 56.10 (2008), pp. 4661–4672.

[PBC14] Jérôme Pasquet, Sandra Bringay, and Marc Chaumont. "Steganalysis with Cover-Source Mismatch and a Small Learning Database". In: *EUSIPCO*. IEEE. EURASIP, 2014, pp. 2425–2429.

[Pen+19] Xingchao Peng et al. "Moment Matching for Multi-Source Domain Adaptation". In: *ICCV*. IEEE, 2019, pp. 1406–1415.

[Pev08] Tomáš Pevný. *Kernel Methods in Steganalysis*. SUNY Binghamton, 2008.

[Pev11] Tomáš Pevný. "Detecting Messages of Unknown Length". In: *MWSF*. Vol. 7880. SPIE. 2011, pp. 300–311.

[PF07] Tomáš Pevný and Jessica Fridrich. "Merging Markov and DCT features for Multi-class JPEG Steganalysis". In: *SSWMC*. Vol. 6505. SPIE. 2007, pp. 28–40.

[PK13] Tomáš Pevný and Andrew Ker. "The Challenges of Rich Features in Universal Steganalysis". In: *MWSF*. Vol. 8665. SPIE. IS&T, 2013, pp. 203–217.

[PN15] Tomáš Pevný and Ivan Nikolaev. "Optimizing Pooling Function for Pooled Steganalysis". In: *WIFS*. IEEE. 2015, pp. 1–6.

[PRM00] Andrea Polesel, Giovanni Ramponi, and V John Mathews. "Image Enhancement via Adaptive Unsharp Masking". In: *TIP* 9.3 (2000), pp. 505–510.

[Ram+05] Rajeev Ramanath et al. "Color Image Processing Pipeline". In: *Signal Processing Magazine* 22.1 (2005), pp. 34–43.

[Rei+19] Stephanie Reinders et al. "Algorithm Mismatch in Spatial Steganalysis". In: *EI* 31 (2019), pp. 1–11.

[Rod+22] Elena Rodríguez-Lois et al. "A Critical Look into Quantization Table Generalization Capabilities of CNN-based Double JPEG Compression Detection". In: *EUSIPCO*. IEEE. 2022, pp. 1022–1026.

[RP20] Alan Ramponi and Barbara Plank. "Neural Unsupervised Domain Adaptation in NLP – Survey". In: *ICCL*. ACL, 2020, pp. 6838–6855.

[RRG19] Tabares-Soto Reinel, Ramos-Pollan Raul, and Isaza Gustavo. "Deep Learning Applied to Steganalysis of Digital Images: A Systematic Review". In: *IEEE Access* 7 (2019), pp. 68970–68990.

[ŠAP22] Dominik Šepák, Lukáš Adam, and Tomáš Pevný. "Formalizing Cover-Source Mismatch as a Robust Optimization". In: *EUSIPCO*. IEEE. EURASIP, 2022.

[Son+15] Xiaofeng Song et al. "Steganalysis of Adaptive JPEG Steganography Using 2D Gabor Filters". In: *IH&MMSec*. 2015, pp. 15–23.

[TRC14] Thanh Hai Thai, Florent Retraint, and Rémi Cogranne. "Statistical Detection of Data Hidden in Least Significant Bits of Clipped Images". In: *Signal Processing* 98 (2014), pp. 263–274. ISSN: 0165-1684.

[Xu+15] Xikai Xu et al. "Robust Steganalysis Based on Training Set Construction and Ensemble Classifiers Weighting". In: *ICIP*. IEEE. 2015, pp. 1498–1502.

[Xue+19] Yiming Xue et al. "A Subspace Learning-Based Method for JPEG Mismatched Steganalysis". In: *MTAP* 78 (2019), pp. 8151–8166.

[Yan+21] Liran Yang et al. "Transfer Subspace Learning based on Structure Preservation for JPEG Image Mismatched Steganalysis". In: *SPIC* 90 (2021), p. 116052.

[Yao+23] Siya Yao et al. "A Survey of Transfer Learning for Machinery Diagnostics and Prognostics". In: *Artificial Intelligence Review* 56.4 (2023), pp. 2871–2922.

[YF20] Yassine Yousfi and Jessica Fridrich. "JPEG Steganalysis Detectors Scalable with respect to Compression Quality". In: *EI* 32 (2020), pp. 1–11.

[YKF18] Yong Yang, Xiangwei Kong, and Chaoyu Feng. "Double-compressed JPEG Images Steganalysis with Transferring Feature". In: *MTAP* 77 (2018), pp. 17993–18005.

[Yu+23] Lifang Yu et al. "RCDD: Contrastive Domain Discrepancy with Reliable Steganalysis Labeling for Cover Source Mismatch". In: *Expert Systems with Applications* (2023), p. 121543.

[Zen+15] Likai Zeng et al. "JPEG Quantization Table Mismatched Steganalysis via Robust Discriminative Feature Transformation". In: *MWSF*. Vol. 9409. SPIE. 2015, pp. 270–278.

[ZH07] Li Zhou and George Hripcsak. "Temporal Reasoning with Medical Data – a Review with Emphasis on Medical Natural Language Processing". In: *JBI* 40.2 (2007), pp. 183–202.

[Zha+18] Zixing Zhang et al. "Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments". In: *TIST* 9.5 (2018), pp. 1–28.

[Zha+19] Xunpeng Zhang et al. "Cover-Source Mismatch in Deep Spatial Steganalysis". In: *WDW*. Springer. 2019, pp. 71–83.

[Zha+21] Lei Zhang et al. "Feature-guided Deep Subdomain Adaptation Network for Dataset Mismatch in Spatial Steganalysis". In: (2021).

[Zha+22] Lei Zhang et al. "Dataset Mismatched Steganalysis using Subdomain Adaptation with Guiding Feature". In: *Telecommunication Systems* 80.2 (2022), pp. 263–276.

[ZQW20] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. "Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey". In: *SSCI*. IEEE, 2020, pp. 737–744.

A Bibliography

Annotated bibliography will be published upon acceptance.