How to augment a small learning set for improving the performances of a CNN-based steganalyzer?
Mehdi Yedroudj, Marc Chaumont, Frédéric Comby

To cite this version:

HAL Id: lirmm-01681883
https://hal-lirmm.ccsd.cnrs.fr/lirmm-01681883v2
Submitted on 3 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
How to augment a small learning set for improving the performances of a CNN-based steganalyzer?

Mehdi YEDROUDJ1,2, Marc CHAUMONT1,3, Frédéric COMBY2
LIRMM1, Univ Montpellier2, CNRS, Univ Nîmes3, Montpellier, France
{mehdi.yedroudj, marc.chaumont, frederic.comby}@lirmm.fr

Abstract

Deep learning and convolutional neural networks (CNN) have been intensively used in many image processing topics during last years. As far as steganalysis is concerned, the use of CNN allows reaching the state-of-the-art results. The performances of such networks often rely on the size of their learning database. An obvious preliminary assumption could be considering that the bigger a database is, the better the results are. However, it appears that cautions have to be taken when increasing the database size if one desire to improve the classification accuracy i.e. enhance the steganalysis efficiency. To our knowledge, no study has been performed on the enrichment impact of a learning database on the steganalysis performance. What kind of images can be added to the initial learning set? What are the sensitive criteria: the camera models used for acquiring the images, the treatments applied to the images, the cameras proportions in the database, etc? This article continues the work carried out in a previous paper in submission [1], and explores the ways to improve the performances of CNN. It aims at studying the effects of base augmentation on the performance of steganalysis using a CNN.

We present the results of this study using various experimental protocols and various databases to define the good practices in base augmentation for steganalysis.

1. Introduction

Convolutional neural networks (CNN) became very popular to solve classification problems in the last five years. Several authors have proposed to use CNNs to solve steganalysis problems [2, 3, 4, 5]. These methods yield encouraging results but remained comparable to the state-of-the-art algorithms performances. Authors have explored many approaches to improve it such as using a phase split [6], an ensemble of CNN [7] the transfer learning [8] or the augmentation of the database [3, 9].

Let us put aside the quest of the best deep learning network architecture for the steganalysis task. In this paper, our objective is to look at a "real-world" problem [10], which is to learn with a small size database. This problem is also known as low regime learning. It is well-known that supervised approaches based on the use of CNNs need a lot of samples when used for steganalysis purposes. The seminal propositions of Qian et al. [2] and Pibire et al. [3] used from 8 000 to 80 000 spatial images resized to $256 \times 256$ (BOSSBase [11] or ImageNet [12]). In 2017 the authors mainly use around 5 000 pairs of images [4, 5, 6, 13], which is probably insufficient. The number of images for the learning has even reached five millions of samples in [9].

In an operational and realistic protocol, the number of available images for the learning task could be much smaller than what is used in "laboratory". Because all the CNN-based steganalysis are sensitive to the cover-source mismatch phenomenon [14, 15, 16], each time the source distribution is modified, the learning process has to be restarted. The aim of this paper is thus to look at the impact of artificial data-augmentation, which is probably more realist than having access to a huge database of a given source distribution. In all cases, using data-augmentation is an automatic process which requires less human time consumption than searching for images of similar distributions.

Today, the classical scenario used to test an embedding algorithm efficiency is to use the BOSSBase [11] for training and testing, assigning 5000 of the 10000 images to the learning database, while the rest used as testing database. A classical way to artificially increase the learning database without changing the labels is to flip and rotate the learning database without interpolation [2].

Recently, Ye et al. [5] proposed to increase the size of the training database, by adding to the initial 50% of BOSSBase, the whole BOWS2 [17] database (this gives a total of 15000 pairs of images for the training set), while the test set is unchanged and is made of the remaining 50% of BOSSBase. This process effectively improves the results in terms of error probability of detection. However, it could be considered as a very lucky measure because the improvement is essentially due to the fact that BOSSBase and BOWS2 share some identical camera models, and a similar "development" process [1].

The question is thus still open: how should we process in order to enrich a learning database? Can we enrich even more the BOSS learning base in order to obtain a huge learning base, and thus improve the steganalysis results? In this paper we intend to experimentally explore efficient ways to increase the learning database of a CNN based steganalyzer. In Section 2, we recall the topology of the CNN used for the various experiments [1]. In Section 3, we describe the experimental protocol and briefly present all the setups. In Section 4, we experimentally explore the different augmentation methods and we draw conclusions on the practical question of the learning database augmentation.

2. Yedroudj-Net CNN

In this paper, our study on the data augmentation for spatial steganalysis is conducted only on the Yedroudj-Net [1]. This CNN

\footnote{The "development" stands for the numerical processes transforming a color RAW image to a $256 \times 256$ 8-bit grey-levels image}
has been created in 2017 and is a mix of the Xu-Net [4] and Ye-Net [5], which are the two best CNNs created up to 2017 for steganalysis purposes. Yedroudj-Net gives better results than Xu-Net [4] and Ye-Net [5] on WOW [18] and S-UNIWARD [19], and also provides better results than an Ensemble Classifier [20] with a Rich Model [21] when compared on a baseline where there is only one CNN, and no tricks such as the use of an ensemble or transfer learning. We have also conducted database augmentation experiments on Xu-Net [4] and Ye-Net [5] and they follow the same trend as Yedroudj-Net.

Yedroudj-Net is composed of a pre-processing block, five convolutional blocks, and a fully connected block made of three fully connected layers followed by a softmax. The network produces a probability distribution over the two class labels: stego or cover image. Fig. 1 illustrates the overall architecture of our CNN.

For more details on Yedroudj-Net, the reader can have a look at the paper [1] and the online code at http://www.lirmm.fr/~chaumont/DemoAndSources.html Note that the hyper-parameters are kept identical.

3.Experimental methodology

3.1.Objectives and Dataset baseline

Our final objective is to increase the size of the learning database of a CNN based steganalysis through data-augmentation in order to improve its performances. Indeed, increasing the number of learning samples is often beneficial for learning efficient features dedicated to a specific task. But, for steganalysis, the samples have to be selected carefully. The "new" samples have to share a "similar distribution" compared to the "original" samples. One thus tries to find distribution-preserving transformations which, when applied on an input cover or precover image, generate synthesized images that follow the same distribution. Those synthesized images could then be integrated into the learning database as additional images in order to increase the CNN classifier efficiency.

In this paper, first, we explore the factors that are influencing a cover distribution such as the camera model, or the development, and second, we propose distribution-preserving transformations that allow to enrich an initial database and to improve the CNN efficiency.

Our baseline setup will thus be working with the BOSSBase split into two sets. We assign 50% of the cover/stego pairs to the "original" training set, and the rest, to the testing set. For the training set, 4000 out of 5000 pairs are randomly selected for training and the remaining 1000 pairs are set aside for validation. Thus, the testing set is made of 5000 pairs. Regardless of the learning database enrichment, the test database will always contain images from and only from BOSSBase. For a fair comparison, we will use the same test base for all the experiments. To summarize, the learning set will always contain at least 4000 pairs of BOSSBase images, and the validation set will always contain 1000 pairs of BOSSBase images.

3.2.Software platform

We used S-UNIWARD [19], and WOW [18], two well-known content-adaptive methods for the embedding in the spatial domain. Note that we used the Matlab implementations (online codes) with the simulator for the embedding and a random key for each embedding. We thus avoid any wrong use of the C++ codes, i.e. a fixed and unique embedding key, as reported in [3].

All experiments were performed with the publicly available Caffe toolbox [22] with necessary modifications, plus digits V5. All tests were run on an NVidia Titan X GPU card.

3.3.Datasets

Due to our GPU computing platform and time limitation, we conduct all the experiments on images of 256 × 256 pixels. To this end, we resampled all the 512 × 512 images to 256 × 256 images, using the imresize() Matlab function with the default parameters (bicubic interpolation with anti-aliasing).

For the various experimental setup, we are using the different databases listed below, and convert them to 256x256 images:

- the BOSSBase v1.01 [11] consisting of 10 000 grey-level images of size 512 × 512, never compressed, and coming from 7 different cameras,
- the BOWS2 [17] consisting of 10 000 grey-level images of size 512 × 512, never compressed, and whose distribution is close to BOSSBase,
- the LIRMMBase [3] consisting of 9 388 grey-level images of size 512 × 512, never compressed, and coming from 7 different cameras. All the used cameras are different from those used in BOSSBase. This database is a variant of the LIRMMBase (http://www.lirmm.fr/~chaumont/LIRMMBase.html) where images with no semantic content have been suppressed. Note that the development (of the RAW images) used in order to obtain the 256×256 images have been done reusing the same script than the one used for generating BOSS and BOWS2 (http://www.lirmm.fr/~chaumont/LIRMMBase/macroProductPGM.sh).
- the PLACES2 [23] containing more than one million of JPEG images coming from unknown cameras. For the experiments, those images are decompressed, converted in grey-level images, and then resized.

For some experiments, we re-run a development process and will use the ImageMagick free and open-source software.

During the CNNs training, we regularly observe the Loss and Accuracy curves, computed on the validation test, to manually stop the training when an over-fitting phenomenon appears. This over-fitting occurs when the Loss curve continues to decrease on the training set but starts to increase on the validation set. For all the experiments, we report the error probability evaluated on the testing set.

http://dde.binghamton.edu/download/
3.4. Description of the different experimental setups

Below, we briefly listed all the experimental setups with a small description explaining each choice:

- **Setup 1: Classical enrichment.** In this setup, the goal is to obtain the performance baseline. The enrichment of the original learning database (made of 4000 pairs) is obtained thanks to the virtual augmentation using the label-preserving flipping and rotations [5], and the enrichment with BOWS2 images. This experiment is presented in Section 4.1.

- **Setup 2: Enrichment with other cameras.** In this setup, the goal is to evaluate the gain/loss of adding images from different cameras from the ones used in the original learning set. This experiment is presented in Section 4.2.

- **Setup 3: Enrichment with strongly dissimilar sources and unbalance proportions.** In this setup, the goal is to evaluate the gain/loss of adding a huge number of images generated using cameras and a development, totally different from those used in the original learning set. This experiment is presented in Section 4.3.

- **Setup 4: Enrichment with the same RAW images but with a different development.** In this setup, the idea is to evaluate the gain/loss of adding the same original RAW images whose development is different from the one used for the original learning set. This experiment is presented in Section 4.4.

- **Setup 5: Enrichment with a re-development of the learning set.** In this setup, the objective is to evaluate the gain/loss of adding the same original images which are re-developed. This experiment is presented in Section 4.5.

4. Results and discussions

4.1. Setup 1: Classical enrichment

In Table 1, we report the error probability obtained when there is no enrichment which means there are 4000 pairs in the learning set (+ 1000 pairs in the validation set), and 5000 pairs in the test set. All images are from BOSSBase. For a cursory comparison, the performance is reported for the Yedroudj-Net, and the Spatial Rich Model + the Ensemble Classifier (SRM + EC), for the embedding algorithm WOW [18] and S-UNIWARD [19] at payload 0.2 bpp.

Yedroudj-Net has an error probability 8% lower for WOW algorithm at 0.2 bpp, and a similar error probability for S-UNIWARD at 0.2 bpp compared to SRM+EC. As reported in [1], Yedroudj-Net obtains similar or better results compared to the state-of-the-art (including versus Xu-Net and Ye-Net) in a fair comparison setup.

<table>
<thead>
<tr>
<th>Steganalysis</th>
<th>Payload</th>
<th>WOW 0.2 bpp</th>
<th>S-UNIWARD 0.2 bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yedroudj-Net</td>
<td>27.8 %</td>
<td>36.7 %</td>
<td></td>
</tr>
<tr>
<td>SRM+EC</td>
<td>36.5 %</td>
<td>36.6 %</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Steganalysis error probability of Yedroudj-Net, and SRM+EC for two embedding algorithms WOW and S-UNIWARD at 0.2 bpp and 0.4 bpp.

Rich Model + the Ensemble Classifier (SRM + EC), for the embedding algorithm WOW [18] and S-UNIWARD [19] at payload 0.2 bpp.

In Table 2, we report the results with no enrichment (noted BOSS), the results with the Virtual Augmentation (VA) of the BOSS’s training set (noted BOSS + VA: Virtual Augmentation consists in label-preserving flipping and rotations), the results with BOWS2 enrichment (noted BOSS + BOWS2), and the results with BOWS2 enrichment + the Virtual Augmentation (noted BOSS+BOWS2+VA). Some of these results have already been given in [1], are re-presented in order to have a self-contained paper. Note that for BOSS+BOWS2, the training set is made of 14 000 pairs (without counting the validation), 32 000 pairs for BOSS+VA (without counting the validation), and for BOSS+BOWS2+VA, the training set is made of 112 000 pairs (without counting the validation).

When the enrichment is obtained by only applying a virtual augmentation (BOSS+VA), a significant improvement is ob-
The decrease of the error probability detection is 3% for WOW (resp. 2% for S-UNIWARD). This enrichment measure was initially proposed in [12] and it is indeed very efficient. The reader should understand that the VA is an easy and low-cost measure in order to significantly improve the performances.

One can also observe better performance when using BOSS+BOWS2 compared to only using BOSSBase. The CNN decreases its detection error probability by 4% for WOW (resp. 2% for S-UNIWARD). As stated in the introduction, BOSSBase and BOWS2 share some identical camera models and a similar "development" process. As also observed in Section 4.2, in a close setup, this enrichment setup ("similar cameras" + "similar development") allows to increase the performances. We guess that in that case, the added images increase the generalization capability of the network.

When the enrichment is obtained with BOSS+BOWS2+VA, again a significant improvement is observed. The decrease of the error probability detection is 7% for WOW (resp. 5% for S-UNIWARD) compared to the no-enrichment setup. Note that the results given in the current Section will be the reference performances for the comparisons given in the next sections.

The observations given in this Section are confirming that if the database augmentation ensures a good diversity of the database, the CNN can improve its detection accuracy. The experiments described in the next sections are thus done in order to better understand the properties that have to be kept when adding images to the original database.

### 4.2. Setup 2: Enrichment with other cameras

<table>
<thead>
<tr>
<th></th>
<th>WOW 0.2 bpp</th>
<th>S-UNIWARD 0.2 bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSS</td>
<td>27.8 %</td>
<td>36.7 %</td>
</tr>
<tr>
<td>BOSS+LIRMM</td>
<td>29.9 %</td>
<td>38.6 %</td>
</tr>
<tr>
<td>BOSS+LIRMM+BOWS2</td>
<td>26.8 %</td>
<td>36.9 %</td>
</tr>
<tr>
<td>BOSS+LIRMM+BOWS2+VA</td>
<td>25.7 %</td>
<td>36.1 %</td>
</tr>
</tbody>
</table>

Table 3: Base Augmentation influence: error probability of Yedroudj-Net, on WOW and S-UNIWARD at 0.2 bpp with a learning base augmented with either LIRMM, LIRMM+BOWS2, or LIRMM+BOWS2+VA.

In Table 3 we report the results with no enrichment (noted BOSS), the results with LIRMM enrichment (noted BOSS + LIRMM), the results with LIRMM and BOWS2 enrichment (noted BOSS + LIRMM + BOWS2), and the results with LIRMM and BOWS2 enrichment + the Virtual Augmentation (noted BOSS + LIRMM + BOWS2 + VA). Note that for BOSS+LIRMM, the training set is made of 14 000 pairs, for BOSS+LIRMM+BOWS2, the training set is made of 23 388 pairs (without counting the validation), and for the BOSS+LIRMM+BOWS2+VA, the training set is made of 187 104 pairs (without counting the validation).

One can observe that results are worst when using BOSS+LIRMM, compared to only using BOSSBase. There is 2% increase of the detection error probabilities for both WOW and S-UNIWARD. For this setup, the enrichment of the learning set is not strongly unbalanced (1 BOSS pair for 2 LIRMM pairs), done with images acquired with different cameras but processed with the same development. It seems that for a beneficial enrichment, the additional images have to be acquired with the same cameras. Additional facts seem to confirm this hypothesis in Section 4.3 and Section 4.4.

When enriching the BOSSBase with BOSS + LIRMM + BOWS2, the results are as good (or a slightly better for WOW) as using the BOSSBase alone. Finally, the results become better when BOSS+BOWS2+LIRMM+VA is used, but the increase in performance is only of 0.9% for S-UNIWARD (resp. 2% for WOW), while using the BOSS+BOWS2 (see Tab.2) give 2% increasing for S-UNIWARD (resp. 4% for WOW).

Those results confirm again that performance is increased if there is an enrichment with images acquired with the same cameras and with the same development (BOWS-2 share similar cameras and a similar development). This tendency seems to contradict the idea that using millions of images, whose distribution is diverse, would be the best solution for increasing the steganalysis results [9]. Indeed, the added images have to share a very similar "distribution" and images have probably to be acquired with the same cameras. In Section 4.3 we explore a little bit more this hypothesis.

### 4.3. Setup 3: Enrichment with strongly dissimilar sources and unbalance proportions

<table>
<thead>
<tr>
<th></th>
<th>WOW 0.2 bpp</th>
<th>S-UNIWARD 0.2 bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSS</td>
<td>27.8 %</td>
<td>36.7 %</td>
</tr>
<tr>
<td>BOSS+PLACES2 1%</td>
<td>34.2 %</td>
<td>41.6 %</td>
</tr>
<tr>
<td>BOSS+PLACES2 10%</td>
<td>40.0 %</td>
<td>43.9 %</td>
</tr>
<tr>
<td>BOSS+PLACES2 100%</td>
<td>44.6 %</td>
<td>45.3 %</td>
</tr>
</tbody>
</table>

Table 4: Base Augmentation influence: error probability of Yedroudj-Net, on WOW and S-UNIWARD at 0.2 bpp with a learning base augmented with different portions of PLACES2.

In Table 4 we report the results with no enrichment (noted BOSS), the results with 1% of PLACES2 enrichment (noted BOSS + PLACES2 1%), the results with 10% of PLACES2 enrichment (noted BOSS + PLACES2 10%), and 1% of PLACES2 enrichment (noted BOSS + PLACES2 100%). Note that for PLACES2 1%, the training set is made of 14 000 pairs (without counting the validation), for PLACES2 10%, the training set is made of 104 000 pairs (without counting the validation), and for the PLACES2 100%, the training set is made of 1 004 000 pairs (without counting the validation).

Whatever the enrichment and whatever the embedding algorithm, the results are always worse than using the BOSSBase alone. For the setup where 1%, resp. 10%, resp. 100% of PLACES2 are added to the learning, the results get worse and worse, with respectively an increase of the detection error for S-UNIWARD (resp. WOW) of 5% (resp. 6%), 7% (resp. 12%), and then 9% (resp. 17%). Note that with an enrichment of 100% of PLACES2 (1 BOSS pair for 251 PLACES2 pairs), the detection is close to a random guessing.
Since the distribution of BOSS and PLACES2 are totally different (PLACES2 results from a JPEG dequantization, and a very diverse set of sources of cameras), the BOSS distribution is lost, and since no re-balancing measures are used during the learning, the BOSS distribution is considered as anecdotal and it is not really taken into account during the learning. Practically, the total loss computed for BOSS images is negligible compared to the total loss computed for PLACES2 images, and thus a minimization of the global loss will mainly concentrate on minimizing the loss associated to the PLACES2 images. Coming back to our previous statement, using millions of images is not sufficient [9], the added images have to share a very similar “distribution” and images have probably to be acquired with the same cameras.

4.4. Setup 4: Enrichment with the same RAW images but with a different development

<table>
<thead>
<tr>
<th></th>
<th>WOW 0.2 bpp</th>
<th>S-UNIWARD 0.2 bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSS</td>
<td>27.8 %</td>
<td>36.7 %</td>
</tr>
<tr>
<td>BOSS+DEV:Res-Bicub</td>
<td>25.7 %</td>
<td>37.5 %</td>
</tr>
<tr>
<td>BOSS+DEV:Res-Spline</td>
<td>26 %</td>
<td>35.8 %</td>
</tr>
<tr>
<td>BOSS+DEV:Res-NoInt</td>
<td>25.6 %</td>
<td>36.2 %</td>
</tr>
<tr>
<td>BOSS+DEV:Res-Crop</td>
<td>34.8 %</td>
<td>44.2 %</td>
</tr>
<tr>
<td>BOSS+BOSS-APL</td>
<td>28.1 %</td>
<td>37.9 %</td>
</tr>
<tr>
<td></td>
<td>26.0 %</td>
<td>35.5 %</td>
</tr>
</tbody>
</table>

Table 5: Base Augmentation influence: error probability of Yedroudj-Net, on WOW and S-UNIWARD at 0.2 bpp with a learning base augmented with different BOSSBase versions.

In Table 4, we report the results with no enrichment (noted BOSS), and the results with 6 different versions of the BOSS-Base, each generated from the RAW images. There is an enrichment with a resizing with a bicubic interpolation (noted BOSS+DEV:Res-Bicub), an enrichment with a resizing with a spline interpolation (noted BOSS+DEV:Res-Spline), the enrichment with a resizing without any interpolation (noted BOSS+DEV:Res-NoInt), the enrichment with no resizing and a central crop (noted BOSS+DEV:Crop), the enrichment with a resizing to a 768 x 768 images without any interpolation and then a central crop (noted BOSS+DEV:Res-Crop), and finally an enrichment with the use of Adobe Photoshop Lightroom 6 instead of ImageMagick, for generating the color images and then resizing to 256 x 256 the images while keeping the width/length ratio (noted BOSS-APL).

From Table 5, we can observe that the enrichment with a crop development (BOSS+DEV:Crop) lead to very bad results. The increase of the detection error of 7% for S-UNIWARD (resp. 7% for WOW). The enrichment with a resize to 768 x 768 followed by a crop (BOSS+DEV:Res+Crop), to a lesser extent, also give bad results with an increase of the detection error of 1% for S-UNIWARD (resp. 0.3% for WOW). Those bad results suggest that a resolution change during the development has a strong impact on the pixels distributions. When looking to the extreme case of the crop development (BOSS+DEV:Crop), we easily understand that the resulting images content change; there is almost no variations and no edges. Thus, an enrichment with a BOSS version whose development does not ensure the same final pixel resolution than BOSS Base will not enrich favourably the learning data-base.

In counterpart, using the same resize procedure with a slight variation on the interpolation (spline, no-interpolation, bicubic), or with the Adobe Photoshop Lightroom Process allows scrounging at most 1% for S-UNIWARD (resp. 2% for WOW). This confirms that additional samples very close to the target BOSS distribution can improve the learning capabilities. Looking back to the various experiment done previously, one can observe that in order to enrich favourably a target database, a favourable measure is to use images acquired with the same cameras than the target database, and to use a very close resizing process than the one used for the target database.

<table>
<thead>
<tr>
<th></th>
<th>WOW 0.2 bpp</th>
<th>S-UNIWARD 0.2 bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSS</td>
<td>27.8 %</td>
<td>36.7 %</td>
</tr>
<tr>
<td>BOSS+all-DEV</td>
<td>23.0 %</td>
<td>33.2 %</td>
</tr>
<tr>
<td>BOSS+BOWS2</td>
<td>23.7 %</td>
<td>34.4 %</td>
</tr>
</tbody>
</table>

Table 6: Base Augmentation influence: error probability of Yedroudj-Net, on WOW and S-UNIWARD at 0.2 bpp with a learning base augmented with different versions of BOSSBase.

In order to push the reflection a little bit more, we made an additional experiment where we regrouped diverse versions of BOSSBase (BOSS+DEV:Res+Bicub, BOSS+DEV:Res+Spline, BOSS+DEV:Res+NoInt, BOSS+DEV:Res+Crop) to the exception of BOSS+DEV:Crop. In Table 6, we report the results with this gathering of various development (noted BOSS+all-DEV), and the results with LIRMM and BOWS2 enrichment (noted LIRMM+BOWS2) and already reported in Section 4.1. Note that for BOSS+all-DEV, the training set is made of 44 000 pairs (without counting the validation), and for LIRMM+BOWS2 the training set is made of 14 000 pairs (without counting the validation).

For those two enrichments, there is a real improvement with a decrease of the error probability of detection of 2-3% for S-UNIWARD (and 4% for WOW). This last result is very interesting and shows that in order to enrich a database, in a practical scenario, there are at least those two options:

Given a target database:

- either Eve (the steganalyst) finds the same camera(s) (used for generating the target database), capture new images, and reproduce the same development than the target database, with a special caution to the resizing.
- either Eve has an access to the original RAW images and reproduce similar developments than the target database with the similar resizing.

The reader should also remember that the Virtual Augmentation is also a good cheap processing measure.

Note that it is unclear which option would be better in a practical case. Additional experiments have to be done in the future. Anyway, those two enrichments show that a very caution process
has to be taken for really improving the results. We believe that those enrichments reduce the over-fitting and also improve the generalization of the learner.

### 4.5 Setup 5: Enrichment with a re-development of the learning set

In all previous setups, given a target database (never compressed 8-bits grey-level $256 \times 256$ images), we were presuming either a prior knowledge of the cameras used for the images acquisitions or a direct access to the RAW versions of the original images. In real-world cases, those knowledges are most of the time not available. Moreover, retrieving the camera models is a very complicated task in a real scenario due to the huge number of cameras.

<table>
<thead>
<tr>
<th></th>
<th>WOW 0.2 bpp</th>
<th>S-UNIWARD 0.2 bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSS</td>
<td>27.8 %</td>
<td>36.7 %</td>
</tr>
<tr>
<td>BOSS+DEV:Translation</td>
<td>34.7 %</td>
<td>47.8 %</td>
</tr>
<tr>
<td>BOSS+DEV:Up-Down-Sampling</td>
<td>31.2 %</td>
<td>42.6 %</td>
</tr>
</tbody>
</table>

Table 7: Base Augmentation influence: error probability of Yedroudj-Net, on WOW and S-UNIWARD at 0.2 bpp with a learning base augmented with a re-development of BOSSBase.

In Table 7 we report the results with no enrichment (noted BOSS), and the results with 2 different redeveloped versions of the BOSSBase, each generated from the original $256 \times 256$ 8-bits grey-level BOSSBase images. The first redevelopment (noted BOSS+DEV:Translation) consists in applying a sub-pixel image translation, of 0.5 pixel, on the padded (symmetric padding) images, and then applying a crop operation to re-obtain a $256 \times 256$ images. The second redevelopment (noted BOSS+DEV:Up-Down-Sampling) consists in applying a Lanczos3 filter for the up-sampling in order to obtain a $512 \times 512$ images, and then downsampling with the same interpolation Kernel to re-obtain images of $256 \times 256$ size. The results are catastrophic with an increase of the error probability of 6% to 11% for S-UNIWARD and 4% to 7% for WOW. The use of a redevelopment does not seem to be a good idea.

### 5. Conclusion

In this paper, we have explored ways to enrich a learning database when steganalysis is done with a CNN. The enrichment is a crucial task since, in the majority of the today’s experiments, the required number of images have to be extremely high due to the huge number of parameters to be learned. Using an insufficient set of examples (images) leads to CNNs that have not "learned enough" and the average efficiency is thus reduced.

After recalling the state-of-the-art of 2017 for the spatial CNN steganalysis, and briefly recalling the state-of-the-art steganalysis approach named Yedroudj-Net, we have presented various results. Additionally to the classical data augmentation which consists to apply flips and rotations on the learning images [12], we observed two others ways for favorably enriching the learning database. The trend is that, in a clairvoyant scenario (knowledge of the embedding algorithm, knowledge of the payload size, approximate knowledge of the of the images distribution), for a given target (test) database, in order to augment its learning database, the steganalyst (Eve) has two choices:

- Either she is able to guess the camera(s) used for generating the target database. She thus captures new images, and reproduce a similar development than the target database, with a special caution to the resizing.
- Either she has an access to the original RAW images and reproduces a similar development than the target database with the similar resizing.

Those two possible ways to enrich the database are very restrictive. As explained in the paper some complementary solutions can be used such as transfer learning [8], or the use of ensembles [7], but the underlying questions of generalizations / cover-source mismatch have to be explored deeper in the future.

### References


Acknowledgments

This work was supported by the University of Montpellier (LIRMM), and the Algerian Ministry of Higher Education / Scientific Research.

Author Biography

Mehdi YEDROUDJ attained his Master’s degree in Computer Science in 2016 from the University of Constantine 2, Algeria. He is currently working toward the Ph.D. degree in the LIRMM laboratory of Montpellier. His research interests are steganography/steganalysis.

Frédéric COMBY received his M.Sc. degree in automatic and micro-electronic systems in 1998, and the Ph.D. degree in automatic and signal processing in 2001 from the University of Montpellier, France. He joined the ICAR Team (image and interaction), in the LIRMM (Laboratory of Informatics, Robotics, and Microelectronics of Montpellier) as Assistant Professor in 2003. His current research topics include image processing, vision and multimedia security.

Marc CHAUMONT received his Engineer Diploma in Computer Sciences at the INSA (National Institute of Applied Sciences) of Rennes, France in 1999, his Ph.D. in Computer Sciences at the IRISA Rennes (INRIA, CNRS, University of Rennes 2, and INSA) in 2003, and his HDR (“Habilitation à Diriger des Recherches”) at the University of Montpellier in 2013. Since September 2005, he is an Assistant Professor in the LIRMM laboratory (Laboratory of Computer Science, Robotics and Micro-electronic) of Montpellier and the University of Nimes. His main research interests are in steganography, steganalysis, digital image forensic, and objects detection with Deep Learning. He is a retired member of the TC of IEEE SPS – Information Forensics and Security, and a reviewer for more than 20 journals (IEEE TIFS, IS&T JEI, ...) and for more than 10 conferences (EI MWSF, IEEE WIFS, ACM IH&MMSec, IEEE ICIP, ...).