Questions & Answers:

" A study on the invariance in security whatever the dimension of images for the steganalysis by deep-learning ", Kévin PLANOLLES, Marc CHAUMONT, Frédéric COMBY, IEEE International Conference on Acoustics, Speech, and Signal Processing, **ICASSP'2023**, 4-10 June 2023, Greek island of Rhodes, Rhodes, 5 pages.

Q1. What are the contributions?

A1. The contribution are:
- the smart crop "2",
- the build of the Nearly-Nested Image Datasets (NNID),
- a proper way to compare networks able to deal with images of multiple sizes,
- the scaling (in order to align the detectability),
- the results showing the coherence of this scaling (using other networks),
- the re-verification that there is not inherent invariance,
- the re-verification that learning with various dimension increases the invariance **without** obtaining a full invariance,
- the observation that "in the case of a lower dimension, there is an overestimation of the number of stegos while in the case of a higher dimension, there is an overestimation of the number of covers (even if a learning is done with multiple dimension)",
- and the proposition of a way to improve the invariance trough the dilated convolution.

Finally, this paper set a potential research axe for finding invariant networks.

Q2. I do not understand how you find Eq. 1 about the Square Root Law. Could you explain?

A2. In the conclusion of the paper [12], it is mentioned that when "using syndrome coding [6] and binary embedding operations (..), it would appear that, fundamentally, steganographic payload capacity,M, is of order √N log N.". So if M = k .√N log N, with N the number of pixels equal to w.h, we have M = k \sqrt(wh) log(wh). The relative payload is thus α = (k/wh) \sqrt(wh) log(wh).

Q3. Why the square root law is not met (when varying the dimension of images)?

A3. The square root law is known to be not fully correct for images. Some hypotheses (asymptotic hypothesis + independencies assumption) of the Square Root Law are not met. Depending on the treatment on images the square root as to be adapted. We can mention for example the paper Steganalysis in Resized Images, J. Kodovský and J. Friddrich, ICASSP'2013. In all the cases the scaling has to be modified and we do it. It does not call in question our proposition to keep the "same difficulty" i.e. the same distribution (histogram) of costs between the dimension.

Q4. Could you explain a little bit more the integral histogram process?

A4. The algorithm consists of building the cumulative histograms for each position (i,j) of the image (of size n.m). Starting from the left-top pixel, each time that we treat the next position, only one bin (compared to the previous histogram) is modified. At the end of the scan, we are obtaining a total of w.h cumulative histograms (one histogram per position). The complexity is thus linear and is O(n.m). Once those w.h cumulative histograms have been computed, the histogram of a rectangular area is deduced using 4 cumulative histograms.

Q5. What is the difference between smart crop [9] and smart crop 2 (defined in the current paper)?

A5. The term "smart crop" was introduced in SID paper [10] (Jessica Fridrich team) in 2018:

"smart crop" was obtained from each large image so that its histogram of local variance (computed from 3 × 3 blocks) was the most similar (in L1 norm) to the histogram of the local variance of the entire 1024 × 1024 image."

In the Alaska paper [2], the smart crop from SID is used:

"the image is cropped following the so-called "smart crop" process defined in [10]"

In our paper, we redefine the notion of the smart crop since we aim to keep the same "difficulty", i.e., preserving the distribution of the steganographic costs. Keeping the local variance is not similar. Additionally, the process to compute the crop, to our knowledge, is new since it uses the histogram integral approach [16]. The process has never been detailed in any steganalysis publication. For avoiding ambiguity, we have renamed our approach as "smart crop 2".

Q6. What is the motivation behind using the Dilated Convolutions?

A6. The inspiration for dilated convolutions comes from [20] and the studies which try to deal with the invariance to the dimension [18] [19]. From a signal point of view, a gaussian filter and a regular subsampling are done when an image is subsampled. The dilated convolutions act as applying the regular subsampling. Knowing that the image data set has various metric dimensions and that in the future we wish to study the resizing it is interesting to include this additional network in the test bed. The Dilated Convolution allows to obtain a gain which is around 0.6% to 1% (see Table 2) and is statistically relevant since "the standard deviation of networks is extremely low; As an example, for a SID trained on 256 × 256 images, it is below 0.02%". This increase in accuracy is indeed showing that using dilated convolutions is a good idea.

Q7. How does the paper posit to other past papers dealing with invariance with to dimension?

A7. Having constant detectability across image dimensions is an important aspect for analyzing the problem finely. Previous papers do not apply this fine evaluation which leads to unclear conclusions. We master the building of the NNID which allows us to decide if or not there is an invariance (past experiments have not addressed the question with this fine granularity that we believe to be necessary for confirming the clues from previous papers). The experiments are thus valuable from a methodological point of view and thus for steganalysis.