Analysis of the Scalability of a Deep-Learning Network for Steganography "Into the Wild"

Hugo RUIZ[1,2], Marc CHAUMONT[1,4], Mehdi YEDROUDJ[1,2], Ahmed OULAD AMARA[1,2], Frédéric COMBY[1,2], Gérard SUBSOL[1,3]
LIRMM[1], Univ Montpellier[2], CNRS[3], Univ Nîmes[4], Montpellier, France

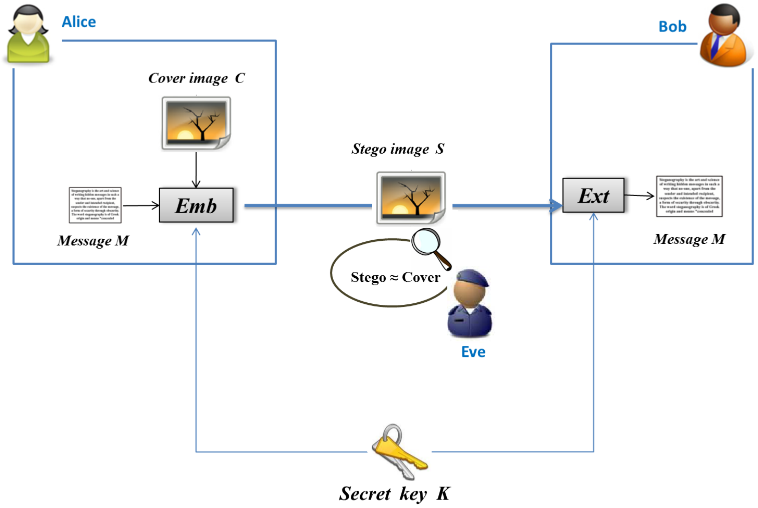
December 29, 2020

ICPR'2021, International Conference on Pattern Recognition,
MMForWILD'2021, Worshop on MultiMedia FORensics in the WILD,
Lecture Notes in Computer Science, LNCS, Springer.

January 10-15, 2021, Virtual Conference due to Covid (formerly Milan, Italy).

Outline

Steganography / Steganalysis



Empirical security measurement:

Steganalysis empirical security measurement ingredients:

- A few state-of-the art **CNN** networks,
- A database,
- A scenario such as the clairvoyant:
  = Laboratory scenario,
  = Worst case attack for Alice.

Empirical security measurement:

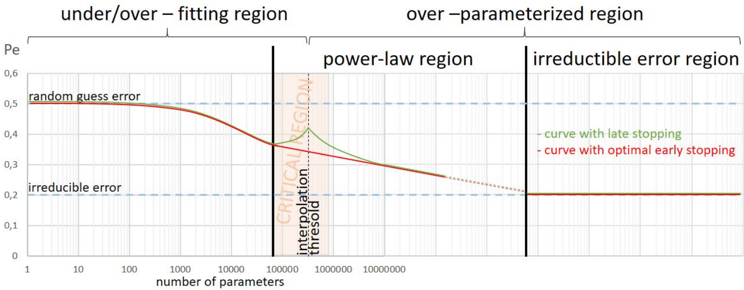
Steganalysis empirical security measurement ingredients:

- A few state-of-the art **CNN** networks,
  $\rightarrow$ Minimum size required?
  - to face to database $\nearrow$,
  - to face to diversity $\nearrow$,
  - to be in the over-parameterized region.

  $\rightarrow$ Accuracy ranking if database is larger?
- A database,
  $\rightarrow$ Minimum size to be better than a random guesser?
  $\rightarrow$ CNNs collapse or not if the training is larger?

(1) Macroscopic black-box first observations:

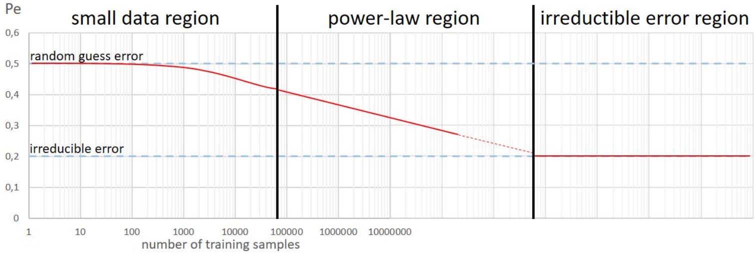
Model scaling **general behavior:**



→ It is beneficial using over-parameterized networks, i.e. with **millions of parameters i.e** $\geq 10^6$.

(2) Macroscopic black-box first observations:

Data scaling **general behavior:**



→ In the power-law region, the more data, the better results,

→ Power-law region seems to start between 10^4 to 10^5 images.

General model for those 2 behaviors:

The test error (noted $\tilde{\epsilon}$) can be simplified[1] in [*]:

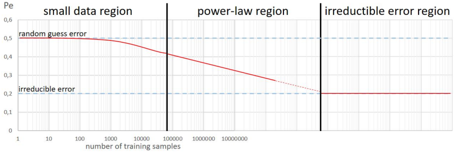
$$\tilde{\epsilon}(m, n) = \underbrace{an^{-\alpha}}_{dataset\ \ power-law} + \underbrace{bm^{-\beta}}_{model\ \ power-law} + c_\infty$$

- a, b, α, β, c_∞ real positive constants,
- n = dataset size, m = model size,
- α and β control the exponential decreasing,
- c_∞ the irreducible error.

[*] Rosenfeld, J.S., Rosenfeld, A., Belinkov, Y., Shavit, N.
A constructive prediction of the generalization error across scales
ICLR'2020, Apr 2020.

[1] in the power-law regions.

Effect of increasing the dataset size:



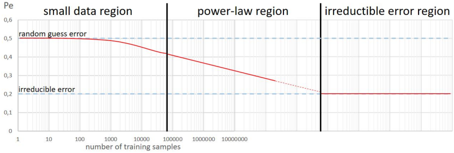
In this paper, we use only one CNN and study the effect of database scaling.

In the **dataset** power-law region,
we should observe the exponential decreasing [*]:

$$\epsilon(n) = a' n^{-\alpha'} + c'_{\infty}$$

[*] Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M.M.A., Yang, Y., Zhou, Y.
Deep Learning Scaling is Predictable, Empirically
Unpublished - ArXiv 1712.00409, 2017.

Why studying the effect of increasing the dataset size?



Why studying this?

- ML community observed this **power-law**. What about **steganalysis?**
- **Database scaling**;

  An important ingredient for **empirical security analysis**?
- **Model scaling** in steganalysis = future work[2].

[2] First observations have been made during JPEG steganalysis Alaska#2 competition, when using the scalable modified EfficientNet network, which is based on the principle of building gradually larger/scalable EfficientNet networks.

Outline

Outline

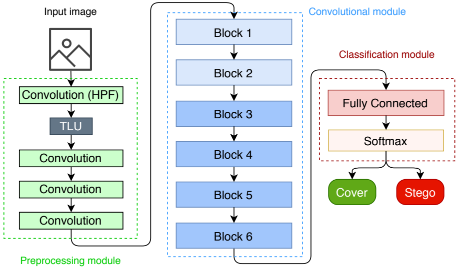
Choice of the network for JPEG steganalysis:

Low Complexity network (LC-Net) [*]:

- One of the state-of-the-art CNN until mid-2020,
- 20 times fewer parameters than SRNet,
- Faster learning than other networks,
- Medium size model (3.10^5 parameters),
  $\rightarrow$ WARNING: model size close to the *interpolation threshold*.
  $\rightarrow$ early stopping during learning.

[*] Huang, J., Ni, J., Wan, L., Yan, J.
A Customized Convolutional Neural Network with Low Model Complexity for JPEG Steganalysis
ACM IH&MMSec'2019. Jul 2019.

LC-Net rapid overview:



Ingredients:

- 30 SRM filters for the pre-processing module,
- 6 blocks using residual connections,
- Blocks 3 to 6 downsample the feature maps,
- ReLU, Batch Norm, and 3x3 convolutions.

Outline

Choice related to the database:

Requirements:

- Grey level images (color steganalysis is not enough understood),
- More than one million images (needs large dataset),
- A controlled database (easier to analysis and generate),
- A diverse database (more realistic),
- A quality factor 75:
  → Robustness to quantization diversity is not enough understood,
  → Will facilitate future comparison with uncontrolled databases;
- Small size images (256×256; memory budget).

The LSSD database is available at:
http://www.lirmm.fr/~chaumont/LSSD.html.

Outline

Choice of the payload:

Objectives:

- Accuracy $\in [60\%, 70\%]$ for a small database ($\simeq$ 20,000 images) i.e. being sufficiently far from the *random-guess* region,
- $\rightarrow$ Large progression margin (when dataset is scaled),
- $\rightarrow$ Room for future works (using better networks).

$\rightarrow$ JUNIWARD at 0.2 bpnzacs for grey-level JPEG 256$\times$256 images from LSSD database with a QF=75.

Outline

Experimental protocol

Essential points:

- 4 learning sets: 20k, 100k, 200k, 1 million (cover+stego) JPEG images,
- 5 models for each learning set (std $<$ to 0.8% for 20k),
- 1 unique test set: 200k (cover+stego) JPEG images,
- LC-Net hyper-parameters are almost the same as the paper,
- Use of an IBM container having access to 2 Tesla V100 GPU.

Outline

Results

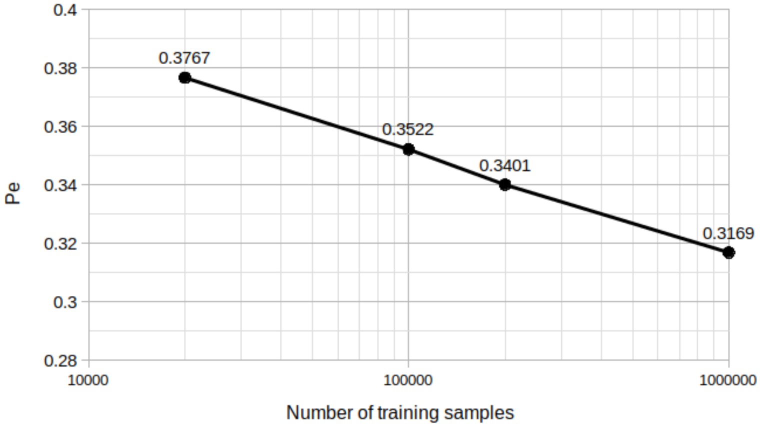


Figure: Average probability of error with respect to the learning database size. Notice that the abscissa scale is logarithmic.

Analysis

Essential points:

- Accuracy improved by 6% from 20k to 1M images,
- LC-Net does not have its performance collapsing,
- Standard deviation is getting smaller and smaller,
  → learning process is more and more stable.

Other facts:

- Time consumption:
  - 20k ≈ 2h
  - 1 million ≈ 10 days
- Memory consumption:
  - 20k ≈ 10 GB (MAT file in double precision)
  - 1 million ≈ 500 GB (MAT file in double precision)

What about power-law?

Using a non-linear regression with Lagrange multipliers:

$$\epsilon(n) = 0.492415n^{-0.086236} + 0.168059$$

- Erroneous to affirm that the irreducible $P_E = 16.8\%$,
- but without much error on the prediction, probability of error for 20M images should be close to **28%**,
- For 2k images it was **37%**,
  $\rightarrow$ 9% increase which is a considerable improvement in steganalysis domain.

Outline

Conclusions (1)

Error power-law is also observed for steganalysis:

- Even with a medium-size model (3×10^5 parameters),
- Even starting with a medium-size database (2×10^4 images).

Take away message:

Increasing a lot (20 million images)

will make you win almost 10% in accuracy

Conclusions (2)

Future work:

- Evaluate with more diversity (quality factors, payload sizes, embedding algorithms, colour, less controlled database),
- Evaluate with other networks,
- Reduce learning time and optimize memory management,
- Find a more precise irreducible error value,
- Study the slope of the power-law depending on the starting point of the CNN (use of transfer, use of curriculum, use of data-augmentation such as pixels-off),
- Find innovative techniques when the database is not huge in order to increase the performances.