Analyse de l'évolutivité d'un réseau d'apprentissage profond pour la stéganalyse d'images

Hugo RUIZ[1], Mehdi YEDROUDJ[1], Marc CHAUMONT[1,2],
Frédéric COMBY[1], Gérard SUBSOL[1]

[1]Research-Team ICAR, LIRMM, Univ. Montpellier, CNRS, France;
[2]Univ. Nîmes, France

hugo.ruiz@lirmm.fr

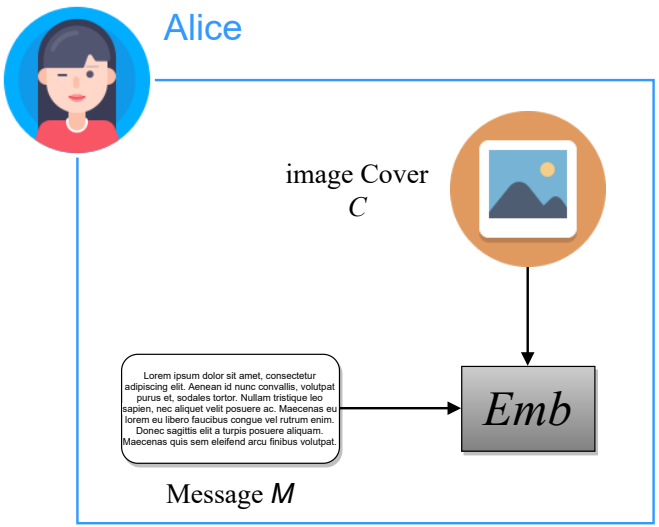
4 novembre 2021

CORESA 2020, COmpression et REpresentation des Signaux Audiovisuels,
3-5 novembre 2021 à Sophia Antipolis, France (reporté à cause de la Covid-19).

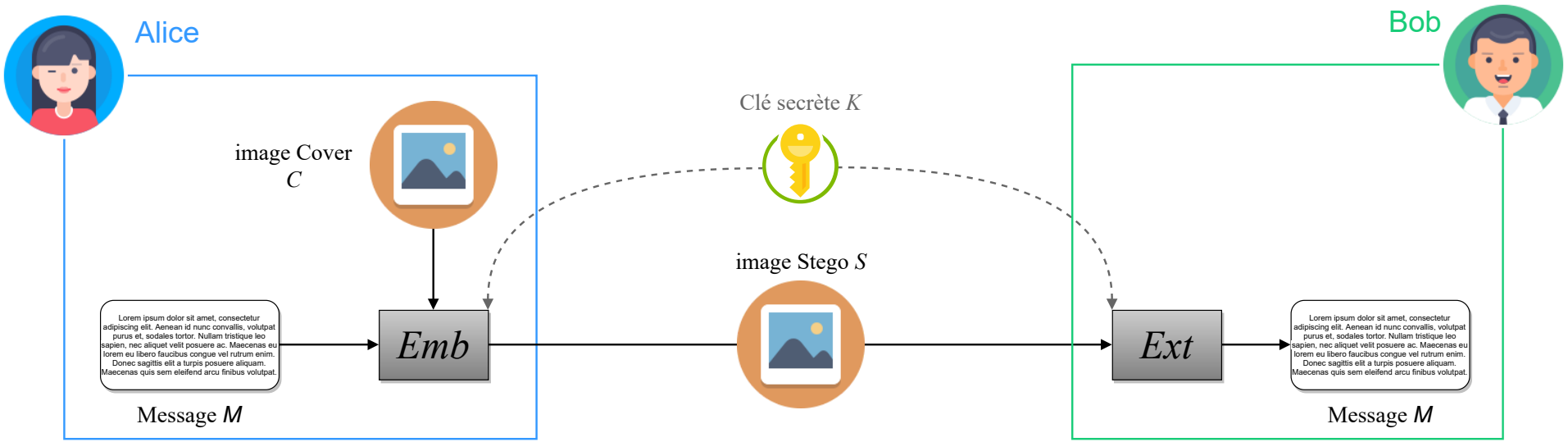
Projet ANR Alaska (ANR-18-ASTR-0009)

Plan

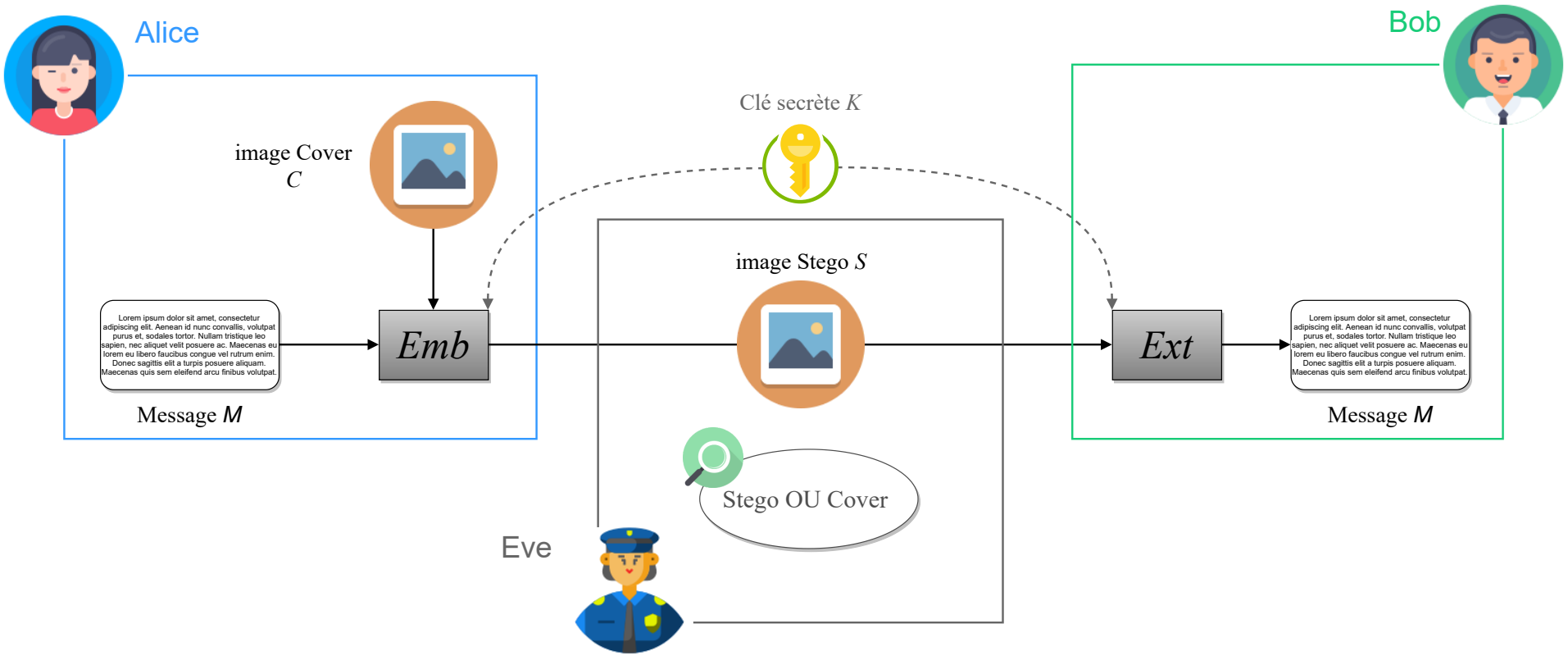
Stéganographie / Stéganalyse



Stéganographie / Stéganalyse



Stéganographie / Stéganalyse



Stéganalyse : détecter les images stego

Les méthodes efficaces :

- fondées sur le *Machine Learning* [KCP13]
- l'apprentissage profond a amélioré les performances [YCC18]
- . . . mais nécessite une base d'apprentissage conséquente

Stéganalyse : détecter les images stego

Éléments pour mesurer empiriquement la sécurité :

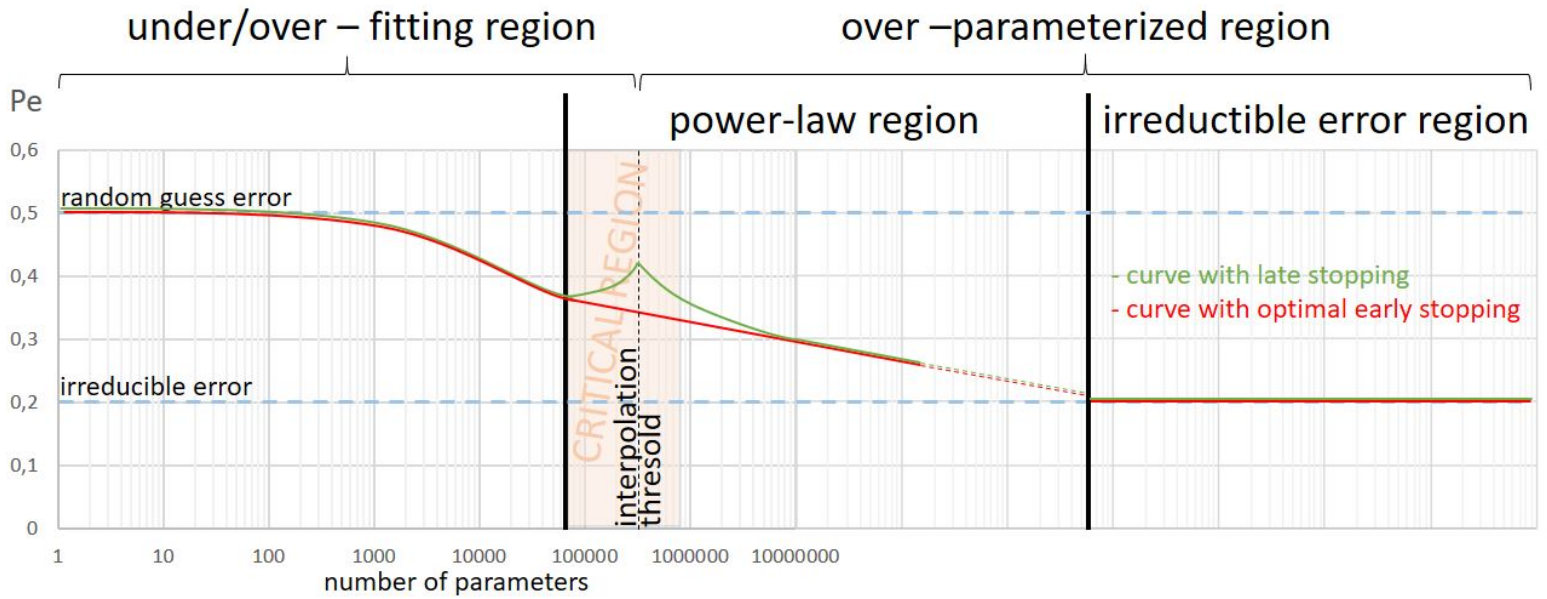
- Un réseau **CNN** (modèle) à l'état de l'art;
- Une base de données (BDD);
- Un scénario tel que le clairvoyant
  = Eve connaît tout sauf la clé d'insertion,
  = pire scénario pour Alice.

Stéganalyse : détecter les images stego

Éléments pour mesurer empiriquement la sécurité :

- Un réseau **CNN** (modèle) à l'état de l'art;
  → Quelle doit être sa taille/grandeur ?
  - face à l'augmentation de la BDD
  - face à l'augmentation de la diversité
- Une base de données (BDD);
  → Taille minimale pour que le réseau soit efficace ?
- Un scénario tel que le clairvoyant
  = Eve connaît tout sauf la clé d'insertion,
  = pire scénario pour Alice.

Évolution théorique d'un point de vue macroscopique (1)



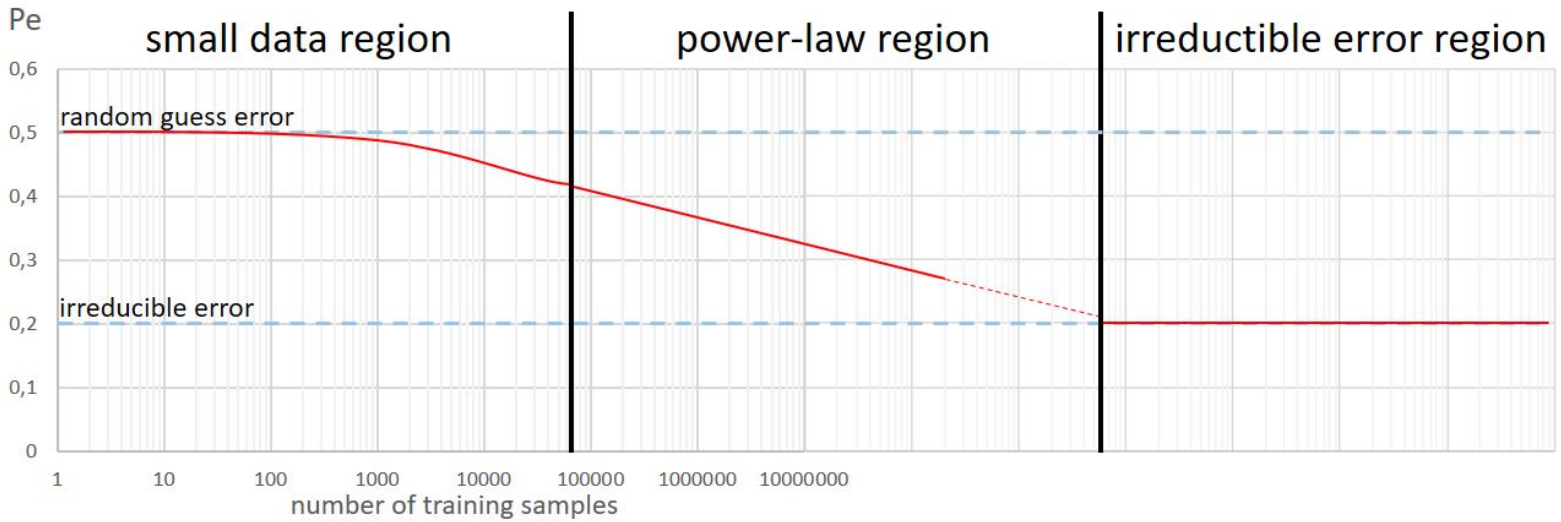
[1]

Figure: Probabilité d'erreur en fonction de la **taille du modèle**

⇒ Les réseaux sur-paramétrés ($> 10^6$) sont intéressants

[1] Inspiré de [HNA+17]

Évolution théorique d'un point de vue macroscopique (2)



[2]

Figure: Probabilité d'erreur en fonction de la **taille de la BDD**

⇒ Plus il y a de données, meilleurs sont les résultats
⇒ La loi de puissance semble être entre 10^4 et 10^5 images

[2] Inspiré de [HNA+17]

Modèle général de ces 2 comportements

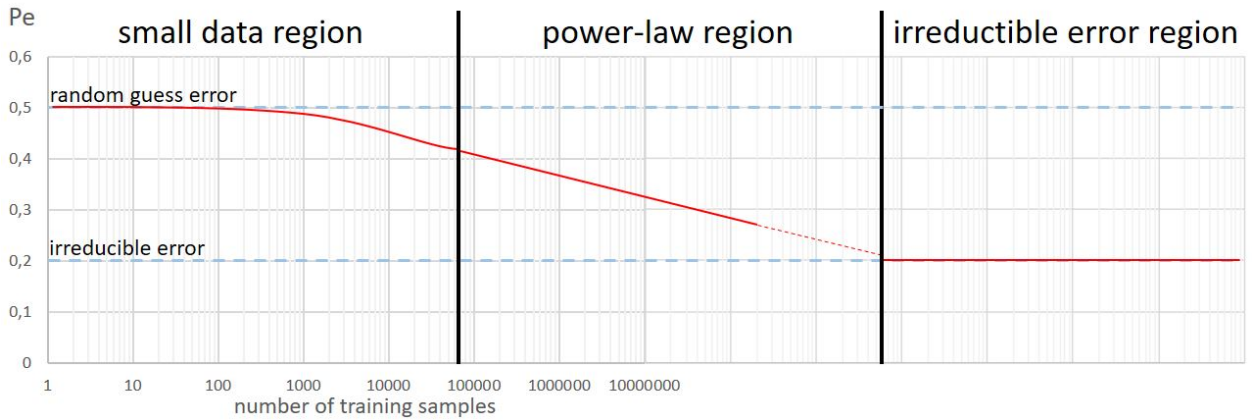
D'après [RRBS20], dans la région de la loi de puissance, la probabilité d'erreur P_e peut se modéliser par l'équation suivante :

$$P_e(m,n) = \underbrace{an^{-\alpha}}_{\text{données}} + \underbrace{bm^{-\beta}}_{\text{modèle}} + c_\infty$$

- a, b, α, β sont des constantes positives;
- m, n respectivement les tailles du modèle et des données;
- c_∞ est une constante appelée erreur irréductible.

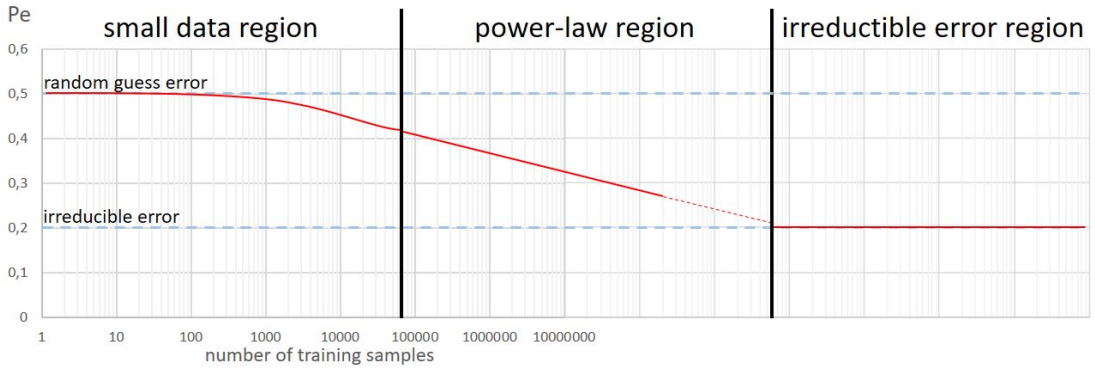
Pourquoi cette étude ?

Étudier l'effet de l'augmentation de la taille de la BDD sur les performances



- Communauté de ML a observé cette **loi**, qu'en est-il pour la **stéganalyse** ?
- Difficile d'analyser la taille du modèle et des données en même temps
- L'**augmentation du jeu de données** : élément important dans l'**analyse empirique en sécurité**.

Modèle simplifié sur la taille des données



Dans ce papier, **un seul CNN est utilisé pour étudier** l'influence de l'augmentation de la taille de la base.

Dans la région de la loi de puissance de la BDD, nous devons observer une courbe décroissante [HNA+17] :

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

Le réseau

Low Complexity network (LC-Net) [HNWY19]

- À l'état de l'art en stéganalyse jusqu'en mi-2020
  Performances peu éloignées de l'état de l'art en 2021.
- 20 fois moins de paramètres que SRNet [BCF19]
  SRNet un réseau à l'état de l'art en 2021.
- Taille moyenne (3.10^5 parameters)
  ⚠ proche du seuil d'interpolation (threshold) ⇒ arrêt manuel
- Apprentissage rapide

La base de données

Utilisation de la **base LSSD** [RYC+21] :

- image : JPEG 256x256
- couleur : non
- facteur de qualité JPEG : 75

- différentes tailles disponibles :
  20k, 100k, 200k, 1M, 2M, 4M *(cover + stego)*

Provenance des images : Alaska#2 (63%), Stego App (19%), BOSS (8%), RAISE (6%), Wesaturate (3%) & Dresden (1%)

La base de données

Utilisation de la **base LSSD** [RYC+21] :

- image : JPEG 256x256
- couleur : non
- facteur de qualité JPEG : 75

- différentes tailles disponibles :
  20k, 100k, 200k, 1M, 2M, 4M *(cover + stego)*

Provenance des images : Alaska#2 (63%), Stego App (19%), BOSS (8%), RAISE (6%), Wesaturate (3%) & Dresden (1%)

Base de test : 200k images issues de LSSD **complètement indépendantes** de la base d'apprentissage.

Comment obtenir les images stego ?

Choix de la méthode d'insertion

"Accuracy" $\in [60\%, 70\%]$ pour une base de petite taille ($\simeq$ 20k images)

- → suffisamment éloignée d'un devineur aléatoire;
- → grande marge de progression lorsque la taille du jeu de données augmente;
- → possibilité d'amélioration avec de meilleurs réseaux.

Accuracy = 1 - P_e

Comment obtenir les images stego ?

Choix de la méthode d'insertion

"Accuracy" $\in [60\%, 70\%]$ pour une base de petite taille ($\simeq$ 20k images)

→ suffisamment éloignée d'un devineur aléatoire;

→ grande marge de progression lorsque la taille du jeu de données augmente;

→ possibilité d'amélioration avec de meilleurs réseaux.

Détails sur l'insertion :

- Algorithme : J-UNIWARD [HF13]
- Charge utile : 0.2 bpnzacs

Accuracy = 1 - P_e

Résultats obtenus

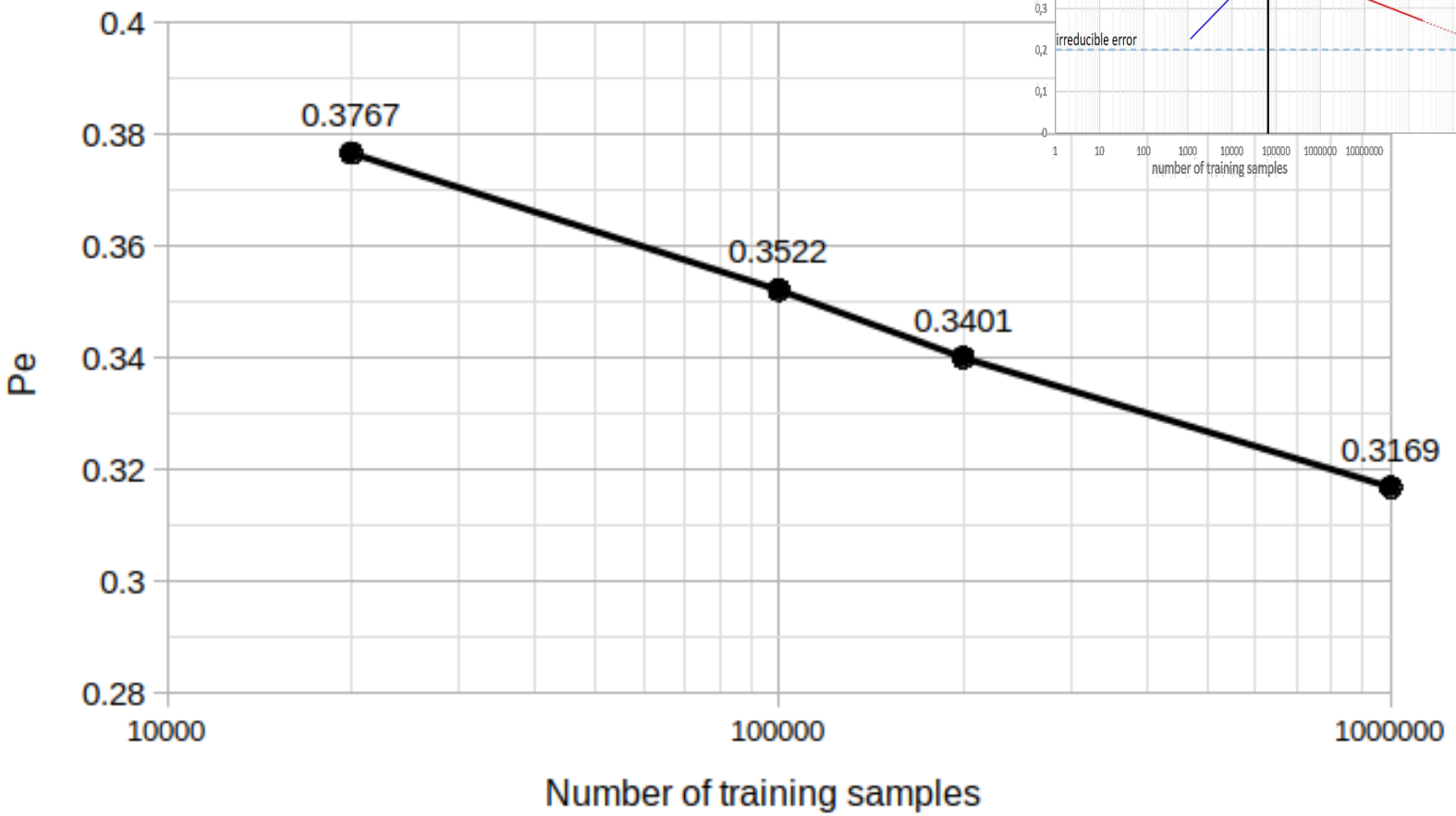


Figure: Probabilité moyenne d'erreur en fonction de la BDD.

Notez que l'échelle des abscisses est logarithmique.

Analyse

Points importants :

- Réduction de P_e de 6% entre 20k et 1M d'images;
- Le réseau LC-Net ne s'effondre pas;
- La variance diminue : l'apprentissage est de plus en plus stable.

Les fichiers décompressés sont au format MAT avec une précision de type `double`.

Analyse

Points importants :

- Réduction de P_e de 6% entre 20k et 1M d'images;
- Le réseau LC-Net ne s'effondre pas;
- La variance diminue : l'apprentissage est de plus en plus stable.

Autres facteurs :

Temps :

- 20k ≈ 2 heures
- 1M ≈ 10 jours

Mémoire :

- 20k ≈ 10 Go
- 1M ≈ 500 Go

Les fichiers décompressés sont au format MAT avec une précision de type `double`.

Conclusions

La loi de puissance proposée de manière générale est bien observée :

- avec un modèle moyen (3.10^5 paramètres);
- en partant d'une base moyenne (2.10^4 images).

Estimation des coefficients : $P'_e(n) = 0.492415 n^{-0.086236} + 0.168059$

Conclusions

La loi de puissance proposée de manière générale est bien observée :

- avec un modèle moyen (3.10^5 paramètres);
- en partant d'une base moyenne (2.10^4 images).

Estimation des coefficients : $P'_e(n) = 0.492415 n^{-0.086236} + 0.168059$

En utilisant une base d'apprentissage de 20 millions d'images permettrait de réduire quasiment 10% la probabilité d'erreur (P_e)

Perspectives pour améliorer l'étude de cette loi

- Rajouter davantage de diversité (facteur de qualité, charge utile, algorithme d'insertion...);
- Changer de réseau;

- Réduire le temps d'apprentissage et optimiser la gestion de mémoire;
- Faire plus d'expériences pour pouvoir calculer la régression non linéaire plus précisément;

- Approfondir l'étude sur la loi de puissance (utilisation de transfert, augmentation de données comme pixels-off [YCC+20])

Remerciements

Merci pour votre attention !

Questions ?

Références I

Mehdi Boroumand, Mo Chen, and Jessica Fridrich.

Deep Residual Network for Steganalysis of Digital Images.

IEEE Transactions on Information Forensics and Security, 14(5):1181 – 1193, May 2019.

Vojtech Holub and Jessica Fridrich.

Digital image steganography using universal distortion.

IH&MMSec 13 Proceedings of the first ACM workshop on Information hiding and multimedia security, 2013(1), 2013.

Joel Hestness, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md. Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou.

Deep Learning Scaling is Predictable, Empirically.

In Unpublished - ArXiv, volume abs/1712.00409, 2017.

Références II

Junwen Huang, Jiangqun Ni, Linhong Wan, and Jingwen Yan.

A customized convolutional neural network with low model complexity for jpeg steganalysis.

In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec'2019, pages 198–203, Paris, France, July 2019.

Sarra Kouider, Marc Chaumont, and William Puech.

Adaptive Steganography by Oracle (ASO).

In Proceeding of the IEEE International Conference on Multimedia and Expo, ICME'2013, pages 1–6, San Jose, California, USA, July 2013.

Jonathan S. Rosenfeld, Amir Rosenfeld, Yonatan Belinkov, and Nir Shavit.

A constructive prediction of the generalization error across scales.

In Proceedings of the Eighth International Conference on Learning Representations, ICLR'2020, April 2020.

Références III

Hugo Ruiz, Mehdi Yedroudj, Marc Chaumont, Frédéric Comby, and Gérard Subsol.

LSSD: a Controlled Large JPEG Image Database for Deep-Learning-based Steganalysis "into the Wild".

In Submited to MultiMedia FORensics in the WILD, MMForWILD'2020, in conjunction with ICPR2020 The 25th International Conference on Pattern Recognition, January 2021.

Mehdi Yedroudj, Frédéric Comby, and Marc Chaumont.

Yedrouj-Net: An Efficient CNN for Spatial Steganalysis.

In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'2018, pages 2092–2096, Calgary, Alberta, Canada, April 2018.

Références IV

Mehdi Yedroudj, Marc Chaumont, Frederic Comby, Ahmed Oulad Amara, and Patrick Bas.

Pixels-off: Data-Augmentation Complementary Solution for Deep-Learning Steganalysis.

In Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security, IHMSec '20, page 39–48, June 2020.