Deep Learning in Steganography and Steganalysis since 2015

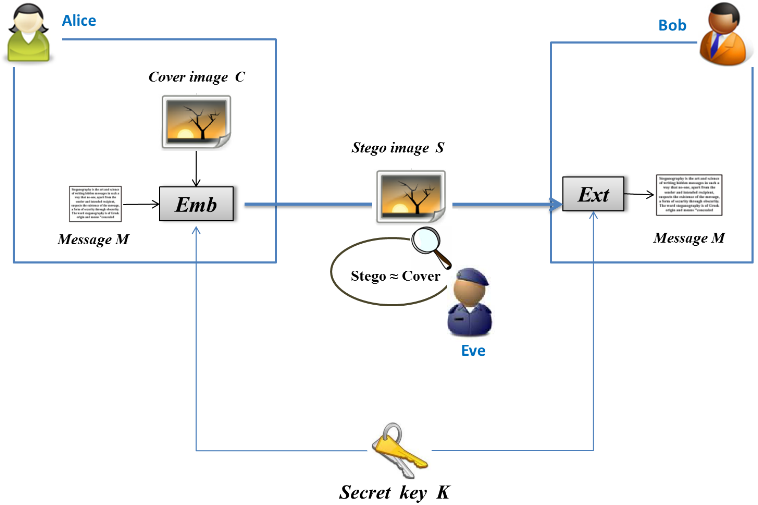
Marc CHAUMONT [1]
(1) LIRMM, Univ Montpellier, CNRS, Univ Nîmes, Montpellier, France

November 2, 2018

Tutorial given at the "Mini - Workshop: Image Signal & Security", Inria Rennes / IRISA. Rennes, France, the 30th of October 2018.

Outline

Steganography / Steganalysis



Embedding example

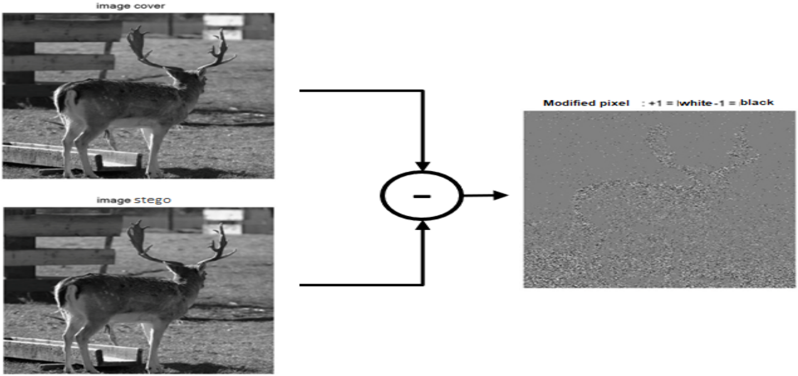
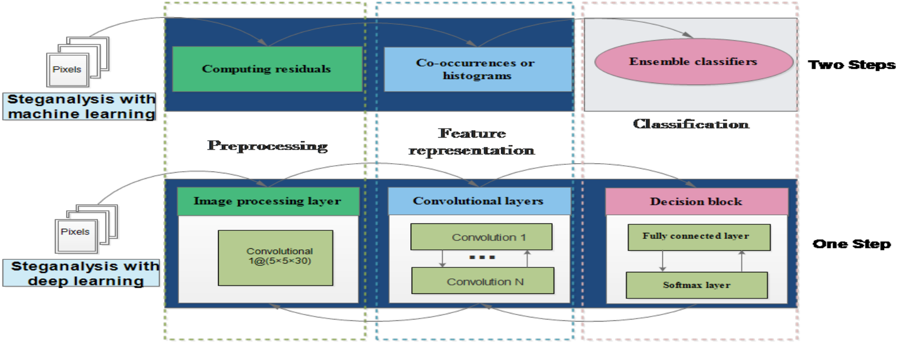


Figure: Example of embedding with S-UNIWARD algorithm (2013) at 0.4 bpp

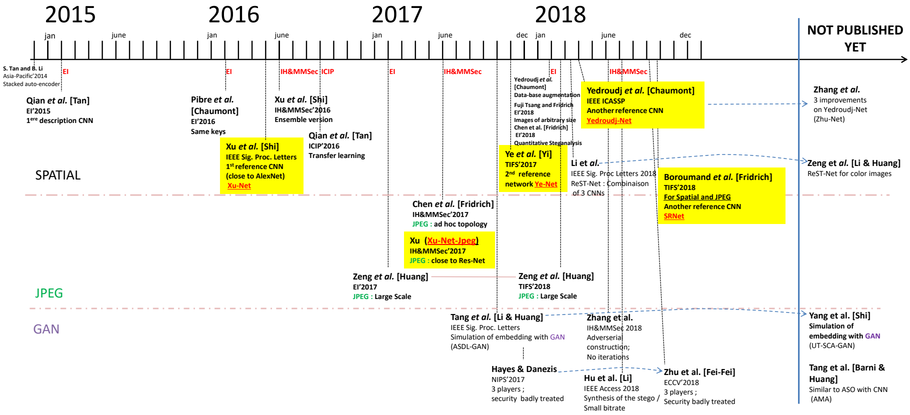
The two families for steganalysis since 2016-2017

- The classic 2-steps learning approach [EC 2012], [Rich 2012] vs. the deep learning approach [Yedroudj-Net 2018], [SRNet 2018]

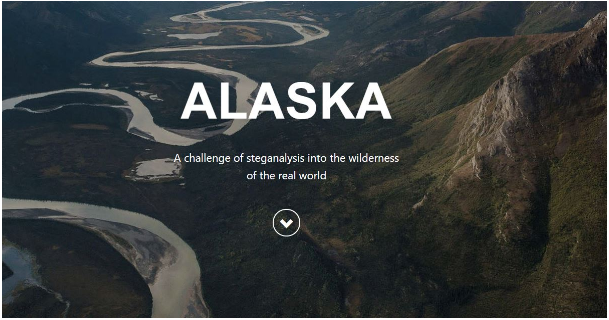


[EC]: "Ensemble Classifiers for Steganalysis of Digital Media", J. Kodovský, J. Fridrich, V. Holub, TIFS'2012
[Rich]: "Rich Models for Steganalysis of Digital Images", J. Fridrich and J. Kodovský, TIFS'2012
[Yedroudj-Net]: "Yedroudj-Net: An Efficient CNN (..)", M. Yedroudj, F. Comby, M. Chaumont, ICASSP'2018
[SRNet] "Deep Residual Network For Steganalysis Of Digital Images", M. Boroumand, Mo Chen, J. Fridrich, TIFS'2018

Chronology


2015
2016
2017
2018
jan
june
jan
june
jan
june
dec
jan
june
dec
NOT PUBLISHED YET
S. Tan and B. Li
Asia-Pacific'2014
Stacked auto-encoder
EI
EI
IH&MMSec
ICIP
EI
IH&MMSec
EI
IH&MMSec
Qian et al. [Tan]
EI'2015
1ere description CNN
Pibre et al. [Chaumont]
EI'2016
Same keys
Xu et al. [Shi]
IH&MMSec'2016
Ensemble version
Xu et al. [Shi]
IEEE Sig. Proc. Letters
1st reference CNN
(close to AlexNet)
Xu-Net
Qian et al. [Tan]
ICIP'2016
Transfer learning
Yedroudj et al.
[Chaumont]
Data-base augmentation
Fuji Tsang and Fridrich
EI'2018
Images of arbitrary size
Chen et al. [Fridrich]
EI'2018
Quantitative Steganalysis
Yedroudj et al. [Chaumont]
IEEE ICASSP
Another reference CNN
Yedroudj-Net
Zhang et al.
3 improvements
on Yedroudj-Net
(Zhu-Net)
Ye et al. [Yi]
TIFS'2017
2nd reference
network Ye-Net
Li et al.
IEEE Sig. Proc Letters 2018
ReST-Net : Combinaison
of 3 CNNs
Zeng et al. [Li & Huang]
ReST-Net for color images
SPATIAL
Boroumand et al. [Fridrich]
TIFS'2018
For Spatial and JPEG
Another reference CNN
SRNet
Chen et al. [Fridrich]
IH&MMSec'2017
JPEG : ad hoc topology
Xu (Xu-Net-Jpeg)
IH&MMSec'2017
JPEG : close to Res-Net
Zeng et al. [Huang]
EI'2017
JPEG : Large Scale
Zeng et al. [Huang]
TIFS'2018
JPEG : Large Scale
JPEG
GAN
Tang et al. [Li & Huang]
IEEE Sig. Proc. Letters
Simulation of embedding with GAN
(ASDL-GAN)
Zhang et al.
IH&MMSec 2018
Adverserial
construction;
No iterations
Yang et al. [Shi]
Simulation of
embedding with GAN
(UT-SCA-GAN)
Hayes & Danezis
NIPS'2017
3 players ;
security badly treated
Hu et al. [Li]
IEEE Access 2018
Synthesis of the stego /
Small bitrate
Zhu et al. [Fei-Fei]
ECCV'2018
3 players ;
Security badly treated
Tang et al. [Barni & Huang]
Similar to ASO with CNN
(AMA)

Alaska



- Challenge from the 05th September 2018 to the 14th March 2019
- Results at IH&MMSec held in Paris in June 2019.
- https://alaska.utt.fr

Outline

An example of a Convolutional Neural Network

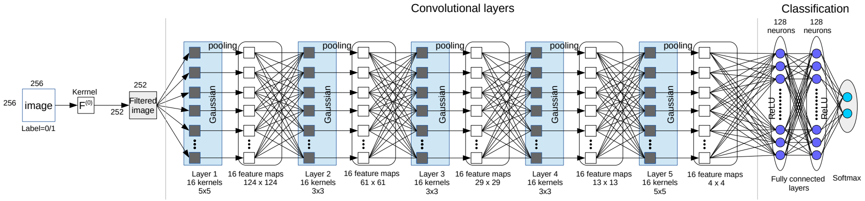


Figure: Qian *et al. 2015* Convolutional Neural Network.

- Inspired by Krizhevsky *et al.*'s CNN 2012,
- Percentage of detection 3 % to 4 % worse than EC + RM.

" ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton, NIPS'2012.

"Deep Learning for Steganalysis via Convolutional Neural Networks," Y. Qian, J. Dong, W. Wang, T. Tan, EI'2015.

Convolution Neural Network: Pre-treatment filter(s)

$$F^{(0)} = \frac{1}{12}\begin{pmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{pmatrix}$$

- CNNs converge more slowly (or not at all?) without preliminary high-pass filter(s)
  - probably true when not much images in the learning set (256x256 at 0.4 bpp less than 10 000 images?),
  - Maybe not so useful when using the cost map?
- Xu-Net, Ye-Net, Yedroudj-Net, Zhu-Net are using a preliminary fixed high-pass filter(s) (eventually updated),
- SRNet learn these filters.

Convolution Neural Network: Layers

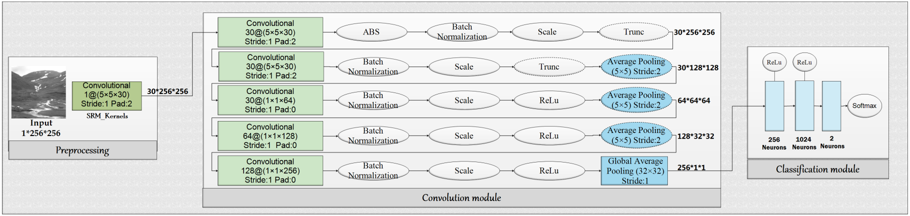


Figure: Yedroudj-Net (2018) Convolutional Neural Network.

In a block, we find these stages:

- A convolution,
- The application of activation function(s),
- A pooling step,
- A normalization step.

"Yedroudj-Net: An Efficient CNN for Spatial Steganalysis", M. Yedroudj, F. Comby, M. Chaumont, ICASSP'2018

Convolution Neural Network: Convolutions

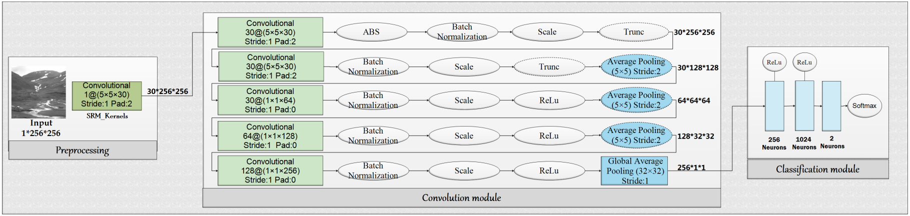


Figure: Yedroudj-Net (2018) Convolutional Neural Network.

$$\tilde{I}_k^{(l)} = \sum_{i=1}^{i=K^{(l-1)}} I_i^{(l-1)} \star F_{k,i}^{(l)},$$

- $I_i^{(l-1)}$: A feature map from the previous Layer,
- $\tilde{I}_k^{(l)}$: Result of the convolution,
- $F_i^{(l)}$: A set of $K^{(l-1)}$ kernel.

"Yedroudj-Net: An Efficient CNN for Spatial Steganalysis", M. Yedroudj, F. Comby, M. Chaumont, ICASSP'2018

Convolution Neural Network: Activation

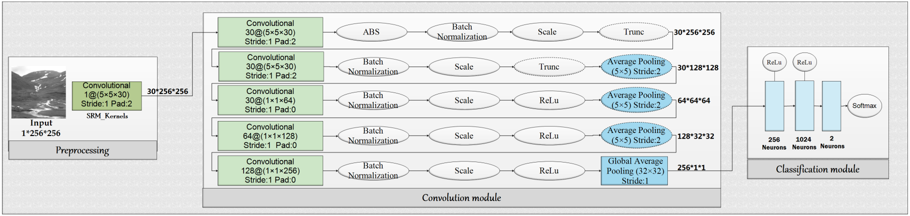


Figure: Yedroudj-Net (2018) Convolutional Neural Network.

Possible activation functions:

- Absolute function: $f(x) = |x|$,
- Sinus function: $f(x) = sinus(x)$,
- Gaussian function (Qian *et al.*'s network) : $f(x) = \frac{e^{-x^2}}{\sigma^2}$,
- ReLU (Rectified Linear Units) : $f(x) = max(0, x)$,
- Hyperbolic tangent: $f(x) = tanh(x)$,
- Truncation (hard tanh parameterized), ..

Convolution Neural Network: Pooling

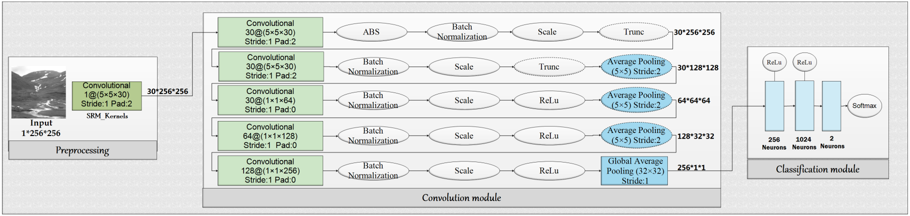


Figure: Yedroudj-Net (2018) Convolutional Neural Network.

Pooling is a local operation computed on a neighborhood:

- local average (preserve the signal),
- or, local maximum (translation invariance property).

\+ a sub-sampling operation.

Convolution Neural Network: Normalization

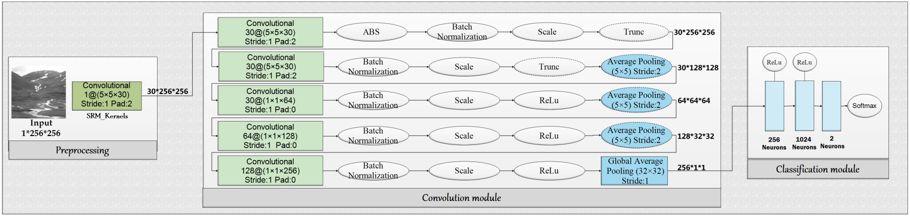
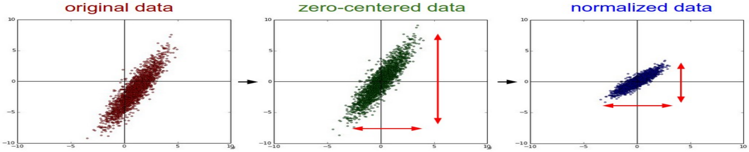


Figure: Yedroudj-Net (2018) Convolutional Neural Network.

Example: Batch Normalization is done on each pixel of a "feature map":

$$BN(X, \gamma, \beta) = \beta + \gamma \frac{X - E[X]}{\sqrt{Var[X] + \epsilon}},$$



Convolution Neural Network: Fully Connected Network

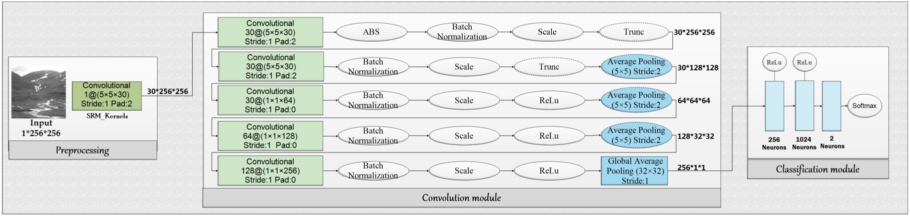
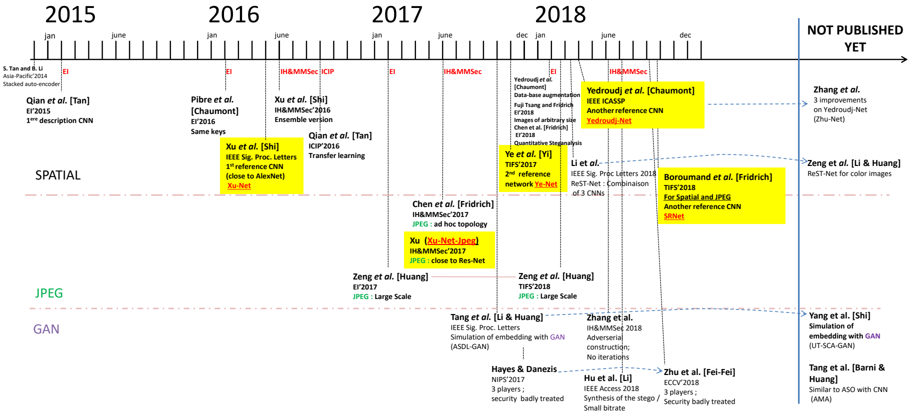


Figure: Yedroudj-Net (2018) Convolutional Neural Network.

- Three layers,
- A softmax function normalizes the values between $[0, 1]$,
- The network issues a value for cover (resp. for stego).

Chronology



2015
2016
2017
2018
jan
june
jan
june
jan
june
dec
jan
june
dec
NOT PUBLISHED YET
S. Tan and B. Li
Asia-Pacific'2014
Stacked auto-encoder
EI
EI
IH&MMSec
ICIP
EI
IH&MMSec
EI
IH&MMSec
Qian et al. [Tan]
EI'2015
1ere description CNN
Pibre et al. [Chaumont]
EI'2016
Same keys
Xu et al. [Shi]
IH&MMSec'2016
Ensemble version
Xu et al. [Shi]
IEEE Sig. Proc. Letters
1st reference CNN
(close to AlexNet)
Xu-Net
Qian et al. [Tan]
ICIP'2016
Transfer learning
Yedroudj et al. [Chaumont]
Data-base augmentation
Fuji Tsang and Fridrich
EI'2018
Images of arbitrary size
Chen et al. [Fridrich]
EI'2018
Quantitative Steganalysis
Ye et al. [Yi]
TIFS'2017
2nd reference network Ye-Net
Yedroudj et al. [Chaumont]
IEEE ICASSP
Another reference CNN
Yedroudj-Net
Li et al.
IEEE Sig. Proc Letters 2018
ReST-Net : Combinaison of 3 CNNs
Boroumand et al. [Fridrich]
TIFS'2018
For Spatial and JPEG
Another reference CNN
SRNet
Zhang et al.
3 improvements on Yedroudj-Net
(Zhu-Net)
Zeng et al. [Li & Huang]
ReST-Net for color images
SPATIAL
Chen et al. [Fridrich]
IH&MMSec'2017
JPEG : ad hoc topology
Xu (Xu-Net-Jpeg)
IH&MMSec'2017
JPEG : close to Res-Net
Zeng et al. [Huang]
EI'2017
JPEG : Large Scale
Zeng et al. [Huang]
TIFS'2018
JPEG : Large Scale
JPEG
GAN
Tang et al. [Li & Huang]
IEEE Sig. Proc. Letters
Simulation of embedding with GAN
(ASDL-GAN)
Zhang et al.
IH&MMSec 2018
Adverserial construction;
No iterations
Yang et al. [Shi]
Simulation of embedding with GAN
(UT-SCA-GAN)
Hayes & Danezis
NIPS'2017
3 players ;
security badly treated
Hu et al. [Li]
IEEE Access 2018
Synthesis of the stego /
Small bitrate
Zhu et al. [Fei-Fei]
ECCV'2018
3 players ;
Security badly treated
Tang et al. [Barni & Huang]
Similar to ASO with CNN
(AMA)

Other "references" networks:

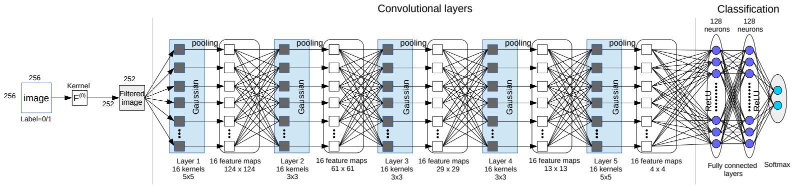


Figure: Qian *et al.* 2015 Convolutional Neural Network.

Xu-Net (may 2016):

Absolute value (first layer),

Activation function: TanH and ReLU,

Normalization function: Batch Normalization (2015),

Ye-Net (nov. 2017):

Filters bank,

Activation function (truncature = "hard tanh"),

8 "layers" and only convolutions,

A version that uses a cost map.

Yedroudj-Net (jan. 2018)

Absolute value

Truncature = "hard tanh"

Batch Normalization

Filters bank

SRNet (sep. 2018):

Filters bank are learned (64)

7 first layers without pooling

Use of shortcuts

Outline

The four families

- **1) Approach by synthesis/no modifications:**
  - Preliminary approaches synthesize a **cover** image [SS-GAN - PCM - Sep 2017], etc.
  - Recent approach synthesize directly a **stego** (with an image generator) [Hu et al -IEEE Access - July 2018]

    $\Rightarrow$ Known to have a low embedding rate + security rely on the generator + must transmit to the extractor
- **2) Approach generating a probability (of modifications) map:**
  - ASDL-GAN [Tang et al. IEEE SPL - Oct 2017], UT-SCA-GAN [Yang et al. ArXiv]

    $\Rightarrow$ only simulations + should test if the "proba" map is usable in practice
- **3) Approach with an adversarial concept (= fooling an oracle = producing adversarial example)**
  - Grandfather are ASO (2012) and MOD (2011)
  - . [Zhang et al. IH&MMSec - June 2018] ; no iteration
  - AMA [Tang et al. - ArXiv] ; only one key ; no equilibrium?
- **4) 3 players approach (equilibrium strategy)**
  - . [Hayes & Danezis - NIPS - Dec 2017], [Zhu et al. - ECCV - Sep 2018]

    $\Rightarrow$ security badly treated for the moment; equilibrium and architecture are hard to find

1) Approach by synthesis

[Hu et al -IEEE Access - July 2018] "A Novel Image Steganography Method via Deep Convolutional Generative Adversarial Networks," in IEEE Access, vol. 6, pp. 38303-38314, 2018.

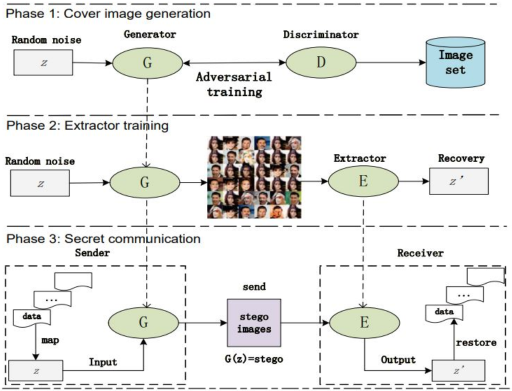


Figure: Steganography Without Embeding (with the use of DCGANs). Figure extracted from the paper [Hu et al. 2018]

The four families

- **1) Approach by synthesis/no modifications:**
  - Preliminary approaches synthesize a **cover** image [SS-GAN - PCM - Sep 2017], etc.
  - Recent approach synthesize directly a **stego** (with an image generator) [Hu et al -IEEE Access - July 2018]

    ⇒ Known to have a low embedding rate + security rely on the generator + must transmit to the extractor
- **2) Approach generating a probability (of modifications) map:**
  - ASDL-GAN [Tang et al. IEEE SPL - Oct 2017], UT-SCA-GAN [Yang et al. ArXiv]

    ⇒ only simulations + should test if the "proba" map is usable in practice
- **3) Approach with an adversarial concept (= fooling an oracle = producing adversarial example)**
  - Grandfather are ASO (2012) and MOD (2011)
  - . [Zhang et al. IH&MMSec - June 2018] ; no iteration
  - AMA [Tang et al. - ArXiv] ; only one key ; no equilibrium?
- **4) 3 players approach (equilibrium strategy)**
  - . [Hayes & Danezis - NIPS - Dec 2017], [Zhu et al. - ECCV - Sep 2018]

    ⇒ security badly treated for the moment; equilibrium and architecture are hard to find

2) Approach generating a probability map

ASDL-GAN [Tang et al. 2017] "Automatic steganographic distortion learning using a generative adversarial network", W. Tang, S. Tan, B. Li, and J. Huang, IEEE Signal Processing Letter, Oct. 2017

UT-SCA-GAN [Yang et al. ArXiv 2018] "Spatial Image Steganography Based on Generative Adversarial Network", Jianhua Yang, Kai Liu, Xiangui Kang, Edward K.Wong, Yun-Qing Shi, ArXiv 2018

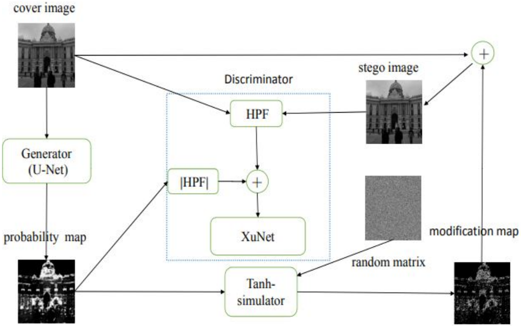


Fig. 1: Steganographic architecture of the proposed UT-SCA-GAN.

Figure: UT-SCA-GAN; Figure extracted from the paper [Yang et al. ArXiv 2018]

The four families

- **1) Approach by synthesis/no modifications:**
  - Preliminary approaches synthesize a **cover** image [SS-GAN - PCM - Sep 2017], etc.
  - Recent approach synthesize directly a **stego** (with an image generator) [Hu et al -IEEE Access - July 2018]

    $\Rightarrow$ Known to have a low embedding rate + security rely on the generator + must transmit to the extractor
- **2) Approach generating a probability (of modifications) map:**
  - ASDL-GAN [Tang et al. IEEE SPL - Oct 2017], UT-SCA-GAN [Yang et al. ArXiv]

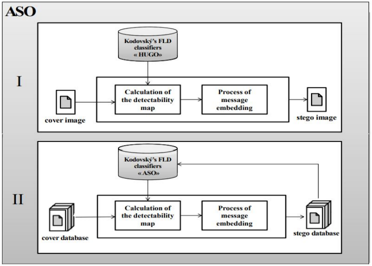
    $\Rightarrow$ only simulations + should test if the "proba" map is usable in practice
- **3) Approach with an adversarial concept (= fooling an oracle = producing adversarial example)**
  - Grandfather are ASO (2012) and MOD (2011)
  - . [Zhang et al. IH&MMSec - June 2018] ; no iteration
  - AMA [Tang et al. - ArXiv] ; only one key ; no equilibrium?
- **4) 3 players approach (equilibrium strategy)**
  - . [Hayes & Danezis - NIPS - Dec 2017], [Zhu et al. - ECCV - Sep 2018]

    $\Rightarrow$ security badly treated for the moment; equilibrium and architecture are hard to find

3) Approach with an adversarial concept

ASO [Kouider et al. ICME'2013] "Adaptive Steganography by Oracle (ASO)", S. Kouider and M. Chaumont and W. Puech, ICME'2013

AMA [Tang et al. ArXiv 2018] "CNN Based Adversarial Embedding with Minimum Alteration for Image Steganography", Weixuan Tang, Bin Li, Shunquan Tan, Mauro Barni, and Jiwu Huang, ArXiv'2018



ASO; Figure from [Kouider et al. ICME'2013]

$$q^{+}_{i,j} = \begin{cases} \rho^{+}_{i,j}/\alpha, & \text{if } \nabla_{z_{i,j}} L(\mathbf{Z}_c, 0; \phi_{\mathcal{C},\mathcal{S}}) < 0, \\ \rho^{+}_{i,j}, & \text{if } \nabla_{z_{i,j}} L(\mathbf{Z}_c, 0; \phi_{\mathcal{C},\mathcal{S}}) = 0, \\ \rho^{+}_{i,j}.\alpha, & \text{if } \nabla_{z_{i,j}} L(\mathbf{Z}_c, 0; \phi_{\mathcal{C},\mathcal{S}}) > 0, \end{cases}$$

$$q^{-}_{i,j} = \begin{cases} \rho^{-}_{i,j}/\alpha, & \text{if } \nabla_{z_{i,j}} L(\mathbf{Z}_c, 0; \phi_{\mathcal{C},\mathcal{S}}) > 0, \\ \rho^{-}_{i,j}, & \text{if } \nabla_{z_{i,j}} L(\mathbf{Z}_c, 0; \phi_{\mathcal{C},\mathcal{S}}) = 0, \\ \rho^{-}_{i,j}.\alpha, & \text{if } \nabla_{z_{i,j}} L(\mathbf{Z}_c, 0; \phi_{\mathcal{C},\mathcal{S}}) < 0, \end{cases}$$

AMA; Eq. from [Tang et al. ArXiv 2018]

The four families

- **1) Approach by synthesis/no modifications:**
  - Preliminary approaches synthesize a **cover** image [SS-GAN - PCM - Sep 2017], etc.
  - Recent approach synthesize directly a **stego** (with an image generator) [Hu et al -IEEE Access - July 2018]

    ⇒ Known to have a low embedding rate + security rely on the generator + must transmit to the extractor
- **2) Approach generating a probability (of modifications) map:**
  - ASDL-GAN [Tang et al. IEEE SPL - Oct 2017], UT-SCA-GAN [Yang et al. ArXiv]

    ⇒ only simulations + should test if the "proba" map is usable in practice
- **3) Approach with an adversarial concept (= fooling an oracle = producing adversarial example)**
  - Grandfather are ASO (2012) and MOD (2011)
  - . [Zhang et al. IH&MMSec - June 2018] ; no iteration
  - AMA [Tang et al. - ArXiv] ; only one key ; no equilibrium?
- **4) 3 players approach (equilibrium strategy)**
  - . [Hayes & Danezis - NIPS - Dec 2017], [Zhu et al. - ECCV - Sep 2018]

    ⇒ security badly treated for the moment; equilibrium and architecture are hard to find

Outline

Conclusion

We saw:

- CNN spatial steganalysis (Yedroudj-Net'2018, Zhu-Net'2019, **SRNet'2018**),
- CNN JPEG steganalysis (JPEG Xu-Net'2017, **SRNet'2018**),
- Performance improvement tricks,
- The GAN families.

What are the hot topics for 2019?

- Alaska challenge, the Cover-Source Mismatch problems, and real life scenarios (whose robust steganography),
- Auto-learnable CNNs, and GANs technology,
- Natural steganography ;-), Batch steganography & Pooled steganalysis.

End of talk

The embedding very rapidly...

More precisely:

- $\mathbf{m} \Longrightarrow \mathbf{c}^*$, such that $\mathbf{c}^*$ is one of the code-word whose syndrome $= \mathbf{m}$, and such that it minimizes the cost function,
- Then, the stego $\leftarrow$ LSB-Matching(cover, $\mathbf{c}^*$).

The STC algorithm is used for coding.

"Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes", T. Filler, J. Judas, J. Fridrich, TIFS'2011.

Performance improvements:

- Virtual Augmentation [Krizhevsky 2012]
- Transfer Learning [Qian et al. 2016] / Curriculum Learning [Ye et al. 2017],
- Using Ensemble [Xu et al. 2016],
- Learn with millions of images? [Zeng et al. 2018],
- Add images from the same cameras and with the similar "development" [Ye et al. 2017], [Yedroudj et al. 2018],
- New networks [Yedroudj et al. 2018], [SRNet 2018], [Zhu-Net - ArXiv], ..
- ...

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton, NIPS'2012,
"Learning and transferring representations for image steganalysis using convolutional neural network", Y. Qian, J. Dong, W. Wang, T. Tan, ICIP'2016,
"Ensemble of CNNs for Steganalysis: An Empirical Study", G. Xu, H.-Z. Wu, Y. Q. Shi, IH&MMSec'16,
"Large-scale jpeg image steganalysis using hybrid deep-learning framework", J. Zeng, S. Tan, B. Li, J. Huang, TIFS'2018,
"Deep Learning Hierarchical Representations for Image Steganalysis," J. Ye, J. Ni, and Y. Yi, TIFS'2017,
"How to augment a small learning set for improving the performances of a CNN-based steganalyzer?", M. Yedroudj, F. Comby, M. Chaumont, EI'2018,
"Yedroudj-Net: An Efficient CNN for Spatial Steganalysis", M. Yedroudj, F. Comby, M. Chaumont, ICASSP'2018,
"Deep Residual Network For Steganalysis Of Digital Images", M. Boroumand, Mo Chen, J. Fridrich, TIFS'2018