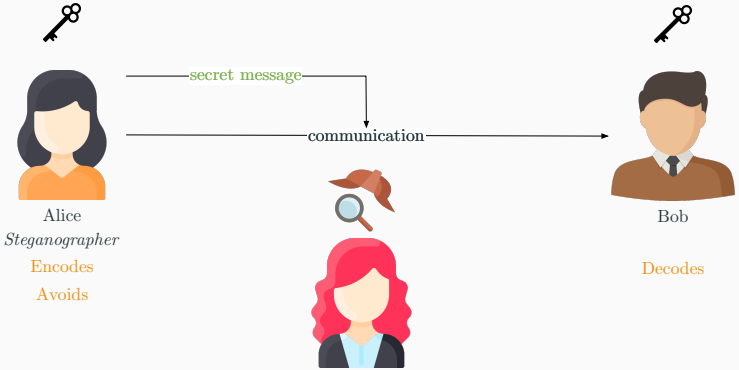
Digital Image Steganography Using Adversarial Embedding

Solène Bernard, **Tomáš Pevný**, Patrick Bas and John Klein

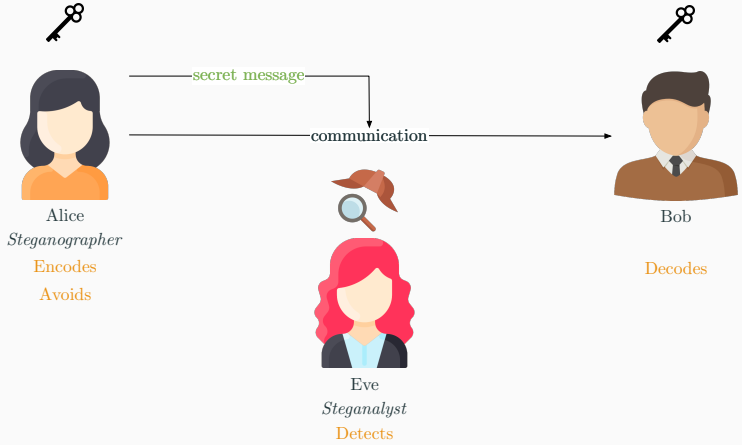
20th November, 2021

Motivation

Context


secret message
communication
Alice
Steganographer
Encodes
Avoids
Eve
Steganalyst
Detects
Bob
Decodes

Context



Kerckhoffs' principle

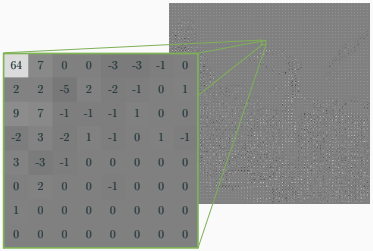
Eve knows everything about Alice's strategy, except the secret key.

Digital grayscale image structure: Spatial or JPEG

Spatial

pixels

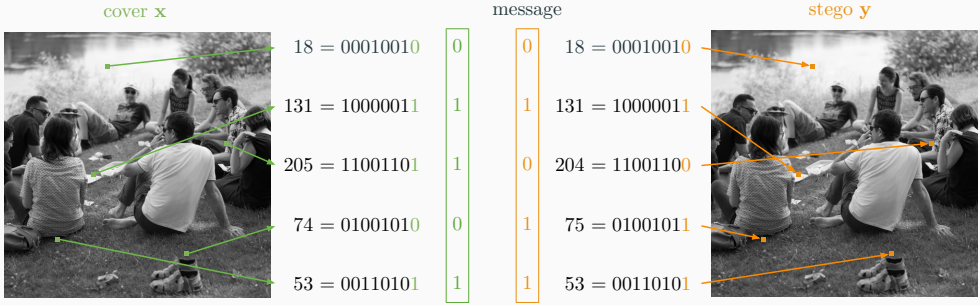
JPEG



DCT coefficients

The first digital images steganographic scheme: LSB replacement

Steganography concept: cover modification. LSB = "Least Significant Bit"

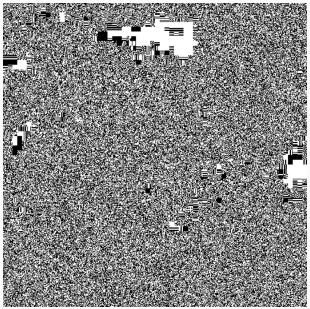


The importance of adaptability to the cover image

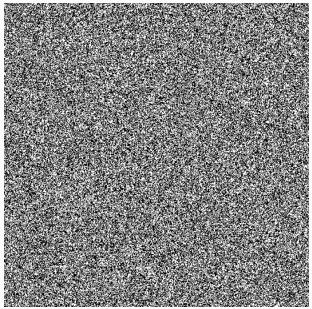
Presence of strong statistical features in *natural* images.

Cover

LSB of Cover



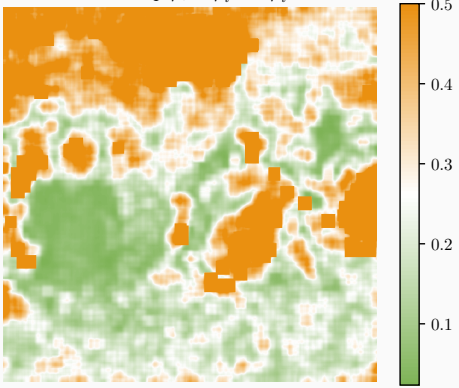
LSB of Stego



Adaptative steganography: cost of modification

Cover

Cost map $\rho_i = \rho_i^{-1} = \rho_i^{+1}$



Alice's objective: embedding while minimizing the distortion

Definition (Distortion)

For an additive cost map $\{\rho_i^b\}$, the distortion D between the cover x and the stego y is equal to:

$$D(\mathrm{x}, \mathrm{y}) = \sum_i \rho_i^{[y_i - x_i]} \tag{1}$$

Algorithm Syndrome-Trellis Code (STC)[1] achieves nearly optimal coding.

[1] Tomáš Filler, Jan Judas, and Jessica Fridrich (2011). "Minimizing additive distortion in steganography using syndrome-trellis codes". In: *IEEE Transactions on Information Forensics and Security* 6.3.

Adaptative steganography: cost of modification

Computation of costs via heuristic principles.

In the spatial domain:

- HUGO[2]
- S-UNIWARD,
- HILL[3],

In the JPEG domain:

- MMx[4]
- J-UNIWARD[5],
- UERD[6].

[2] T. Pevny, T. Filler, and P. Bas (2010). "Using High-Dimensional Image Models to Perform Highly Undetectable Steganography". In: *Information Hiding 2010*. event-place: Calgary, Canada.

[3] Bin Li et al. (2014). "A new cost function for spatial image steganography". In: *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 4206–4210.

[4] Younhee Kim, Zoran Duric, and Dana Richards (2006). "Modified matrix encoding technique for minimal distortion steganography". In: *International Workshop on Information Hiding*. Springer, pp. 314–327.

[5] Vojtěch Holub, Jessica Fridrich, and Tomáš Denemark (2014). "Universal distortion function for steganography in an arbitrary domain". In: *EURASIP Journal on Information Security* 2014.1, pp. 1–13.

[6] Linjie Guo et al. (2015). "Using statistical image model for JPEG steganography: Uniform embedding revisited". In: *IEEE Transactions on Information Forensics and Security* 10.12, pp. 2669–2680.

Simulated steganography - The theory

Simulation of embedding, with

- an additive cost map $\{\rho_i^b\}$,
- a length of the hypothetical message $|m|$.

Simulated steganography - The theory

Simulation of embedding, with

- an additive cost map $\{\rho_i^b\}$,
- a length of the hypothetical message $|m|$.

Categorical distribution of change for each image coefficient:

$$\pi_i^b = P_i^j(\rho_i, \lambda) = \frac{e^{-\lambda \rho_i^b}}{\sum_{b' \in \{-1,0,+1\}} e^{-\lambda \rho_i^{b'}}}, \tag{2}$$

Simulated steganography - The theory

Simulation of embedding, with

- an additive cost map $\{\rho_i^b\}$,
- a length of the hypothetical message $|m|$.

Categorical distribution of change for each image coefficient:

$$\pi_i^b = P_i^j(\rho_i, \lambda) = \frac{e^{-\lambda \rho_i^b}}{\sum_{b' \in \{-1,0,+1\}} e^{-\lambda \rho_i^{b'}}}, \tag{2}$$

where λ determined from the entropy constraint:

$$\text{Entropy}(\{\pi_i^b\}) = H(\{\pi_i^b\}) = |m|, \tag{3}$$

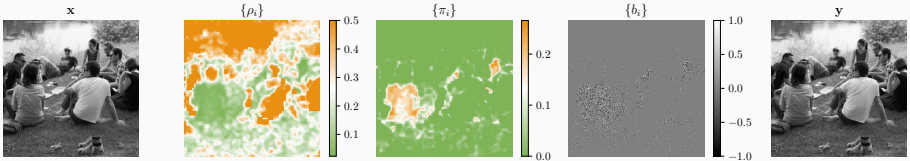
with H the binary entropy.

Simulated steganography

Pipeline from the cover to the simulated stego:

$$\mathsf{x} \longrightarrow \{\rho_i^b\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow{\text{draw}} \{b_i\}_i \longrightarrow \mathsf{y} = \mathsf{x} + \mathsf{b}$$

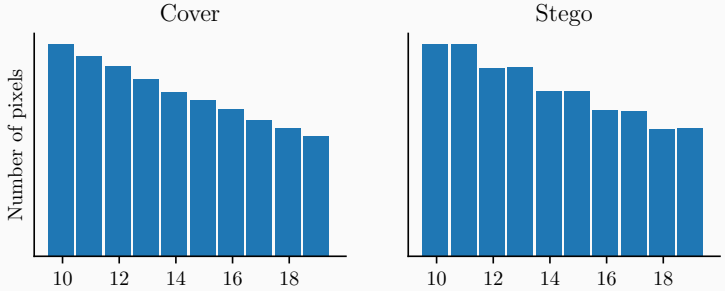
cover | costs | probabilities | changes | stego



Steganalysis

Steganalysis

Historically (until 2015): via features extraction followed by classification, for example histogram attack[7].



[7] Jessica Fridrich, Miroslav Goljan, and Rui Du (2001). "Detecting LSB steganography in color, and gray-scale images". In: *IEEE multimedia* 8.4, pp. 22–28.

Steganalysis

Since 2015: Convolution Neural Networks (CNN).

Classifier f, where $f : \mathcal{I} \to [0, 1]$.

$$\begin{cases} \text{If } f(x) < 0.5 & \text{then x classified as cover} \\ \text{If } f(x) \geq 0.5 & \text{then x classified as stego} \end{cases}$$

State-of-the-art of architectures:

- XU-Net[8],
- SRNet[9],
- Efficient-Net[10].

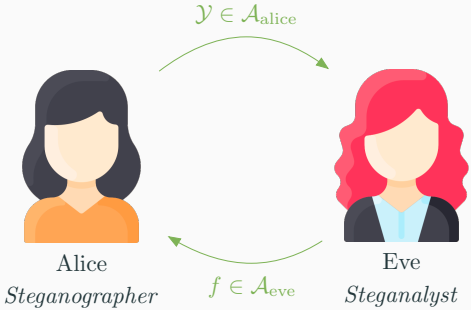
[8] Guanshuo Xu (2017). "Deep convolutional neural network to detect J-UNIWARD". In: *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, pp. 67–73.

[9] Mehdi Boroumand, Mo Chen, and Jessica Fridrich (2018). "Deep residual network for steganalysis of digital images". In: *IEEE Transactions on Information Forensics and Security* 14.5, pp. 1181–1193.

[10] Mingxing Tan and Quoc Le (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *International Conference on Machine Learning*. PMLR, pp. 6105–6114.

Staganography as a game

Staganography as a game



- $\mathcal{A}_{\text{alice}}$: set of actions of Alice
- $\mathcal{A}_{\text{eve}}$: set of actions of Eve

“Solving” the game a minmax protocol

Definition (Eve's utility)

For $(\mathcal{Y}, f) \in \mathcal{A}_a \times \mathcal{A}_e$, Eve's utility is the accuracy of classification of f between $\mathcal{X}$ and $\mathcal{Y}$, ie:

$$u_e(\mathcal{Y}, f) = \frac{1}{2}\left(\mathbb{E}_{x \sim P_{\mathcal{X}}}[f(x) < 0.5] + \mathbb{E}_{y \sim \mathcal{Y}}[f(y) \geq 0.5]\right) \tag{4}$$

Players with antagonists roles

Definition (Eve's utility)

For $(\mathcal{Y}, f) \in \mathcal{A}_a \times \mathcal{A}_e$, Eve's utility is the accuracy of classification of f between $\mathcal{X}$ and $\mathcal{Y}$, ie:

$$u_e(\mathcal{Y}, f) = \frac{1}{2}\left(\mathbb{E}_{\mathrm{x}\sim P_{\mathcal{X}}}[f(\mathrm{x}) < 0.5] + \mathbb{E}_{\mathrm{y}\sim \mathcal{Y}}[f(\mathrm{y}) \geq 0.5]\right) \tag{4}$$

Definition (Alice's utility)

For $(\mathcal{Y}, f) \in \mathcal{A}_a \times \mathcal{A}_e$, Alice's utility is equal to the opposite of Eve's utility:

$$u_a(\mathcal{Y}, f) = -u_e(\mathcal{Y}, f) \tag{5}$$

Game theoretical definition of the problem

Solution concept:

$$\arg\min_{\mathcal{A}_a} \max_{\mathcal{A}_e} u_e(\mathcal{Y}, f)$$

Approximation using double oracle algorithm

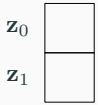
$\mathbf{z}_0$ 

Alice's actions $\mathcal{A}_a$

f_0

Eve's actions $\mathcal{A}_e$

Approximation using double oracle algorithm

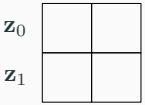


Alice's actions $\mathcal{A}_a$

f_0

Eve's actions $\mathcal{A}_e$

Approximation using double oracle algorithm

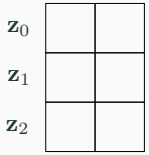


Alice's actions $\mathcal{A}_a$

$f_0 \quad f_1$

Eve's actions $\mathcal{A}_e$

Approximation using double oracle algorithm

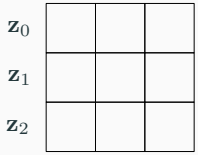


Alice's actions $\mathcal{A}_a$

$f_0 \quad f_1$

Eve's actions $\mathcal{A}_e$

Approximation using double oracle algorithm

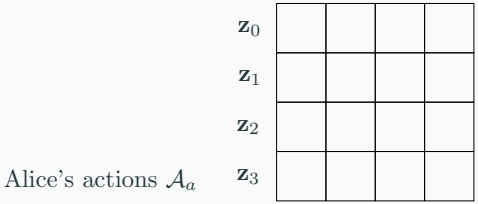


Alice's actions $\mathcal{A}_a$

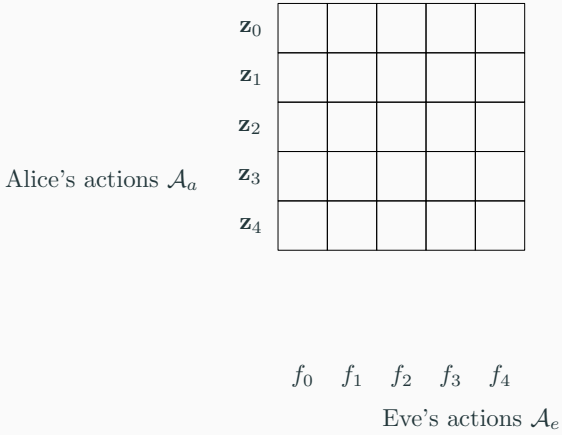
$f_0 \quad f_1 \quad f_2$

Eve's actions $\mathcal{A}_e$

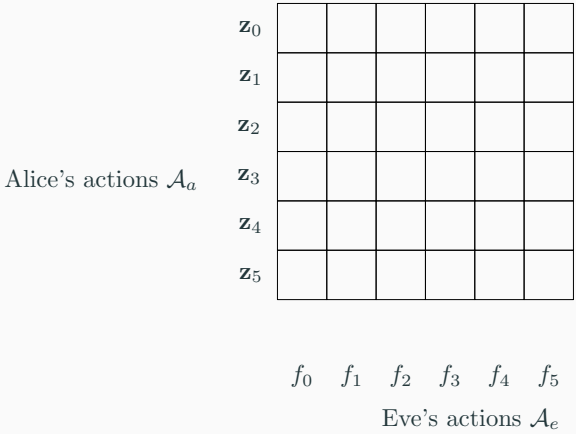
Approximation using double oracle algorithm



Approximation using double oracle algorithm



Approximation using double oracle algorithm



Initialization: $\mathcal{A}_e = \mathcal{F}^0 = \{f^0\}$, $\mathcal{A}_a = \{\mathcal{Y}^0\}$.

Initialization: $\mathcal{A}_e = \mathcal{F}^0 = \{f^0\}$, $\mathcal{A}_a = \{\mathcal{Y}^0\}$.
At k^{th} iteration, the two following macro-steps:

1. Alice's turn.

$$y^* = \arg\min_{y \in \mathcal{A}_a} \max_{f \in \mathcal{F}^{k-1}} f(y). \tag{6}$$

Initialization: $\mathcal{A}_e = \mathcal{F}^0 = \{f^0\}$, $\mathcal{A}_a = \{\mathcal{Y}^0\}$.
At k^{th} iteration, the two following macro-steps:

1. Alice's turn.
$$y^* = \arg\min_{y \in \mathcal{A}_a} \max_{f \in \mathcal{F}^{k-1}} f(y). \tag{6}$$

2. Eve's turn.
Creation of a new detector f^k and appends it to the pool $\mathcal{F}^{k-1}$, i.e. $\mathcal{F}^k = \mathcal{F}^{k-1} \cup \{f^k\}$.

Details of player's actions

1. Alice's new action:

$$\mathsf{x} \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow{\text{draw}} \{b_i\}_i \longrightarrow \mathsf{y} \underset{\nabla_{\mathsf{y}} f(\mathsf{y})}{\longrightarrow} f(\mathsf{y})$$

ADV-EMB[11] proposes the following update rule, with $\alpha = 2$:

$$\rho_i^{+,new} = \begin{cases} \rho_i^+ / \alpha & \text{if } \frac{\partial f}{\partial y_i}(\mathsf{y}) < 0, \\ \rho_i^+ & \text{if } \frac{\partial f}{\partial y_i}(\mathsf{y}) = 0, \\ \rho_i^+ \alpha & \text{if } \frac{\partial f}{\partial y_i}(\mathsf{y}) > 0. \end{cases} \tag{7}$$

[11] Weixuan Tang et al. (2019). "CNN-based Adversarial Embedding for Image Steganography". In: *IEEE Transactions on Information Forensics and Security*.

Details of player's actions

1. Alice's new action:

$$\mathsf{x} \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow{\text{draw}} \{b_i\}_i \longrightarrow \mathsf{y} \underset{\xleftarrow{\nabla_{\mathsf{y}} f(\mathsf{y})}}{\longrightarrow} f(\mathsf{y})$$

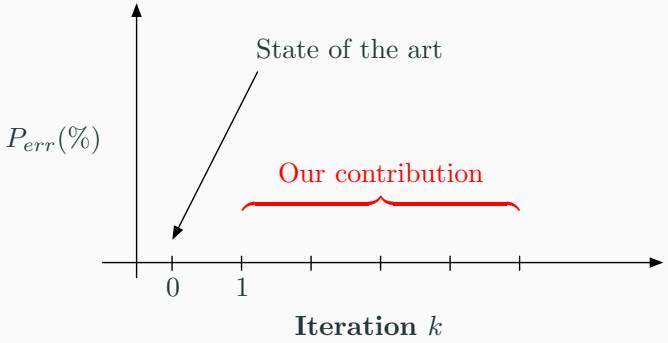
ADV-EMB[11] proposes the following update rule, with $\alpha = 2$:

$$\rho_i^{+,new} = \begin{cases} \rho_i^+ / \alpha & \text{if } \frac{\partial f}{\partial y_i}(\mathsf{y}) < 0, \\ \rho_i^+ & \text{if } \frac{\partial f}{\partial y_i}(\mathsf{y}) = 0, \\ \rho_i^+ \alpha & \text{if } \frac{\partial f}{\partial y_i}(\mathsf{y}) > 0. \end{cases} \tag{7}$$

2. Eve's new action: training of a new classifier f^k to discriminate between $\mathcal{X}$ and $\mathcal{Y}^k$.

[11] Weixuan Tang et al. (2019). "CNN-based Adversarial Embedding for Image Steganography". In: *IEEE Transactions on Information Forensics and Security*.

Results - How to read the plots

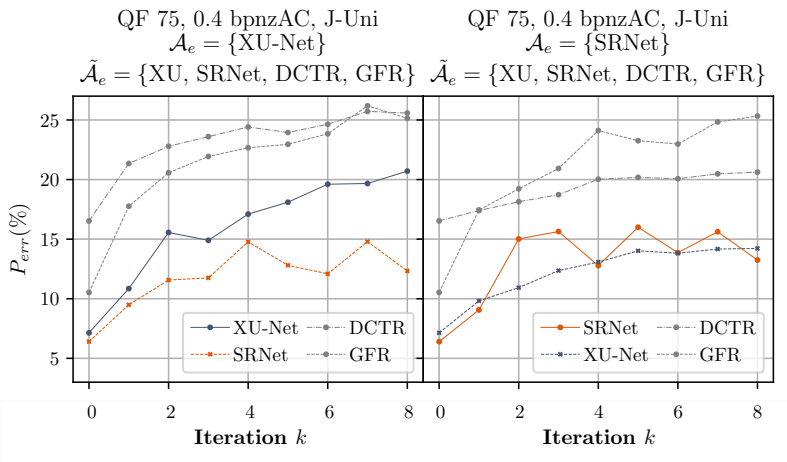


Results - Experimental setup

Mismatch between what assumes Alice can differ from the real actions of Eve.

$$\underbrace{\mathcal{A}_{\text{eve}}}_{\text{Assumed Eve's actions}} \neq \underbrace{\tilde{\mathcal{A}}_{\text{eve}}}_{\text{Real Eve's actions}}$$

Results - With different adversaries for JPEG QF 75, payload 0.4 bpnzAC



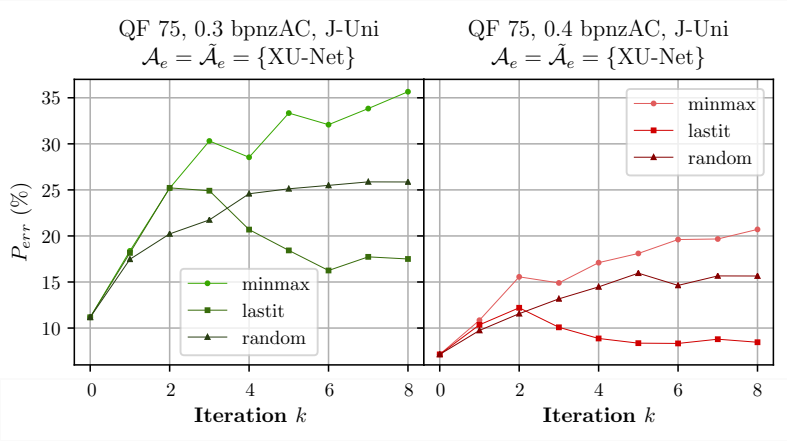
Comparison of different strategies

Comparison of different strategies:

- Minmax: $y^* = \arg\min_{y \in \mathcal{A}_a} \max_{f \in \mathcal{F}^{k-1}} f(y)$;
- Last iteration[12]: $y^* = \arg\min_{y \in \mathcal{A}_a} f^{k-1}(y)$;
- Random: $y^* = \arg\min_{y \in \mathcal{A}_a} f^i(y)$ for $i \sim U(\{0, \dots, k-1\})$.

[12] Weixuan Tang et al. (2019). "CNN-based Adversarial Embedding for Image Steganography". In: *IEEE Transactions on Information Forensics and Security*.

Results - Comparison of strategies for JPEG at QF 75 and different payloads (bpnzAC)



Flaws of ADV-EMB

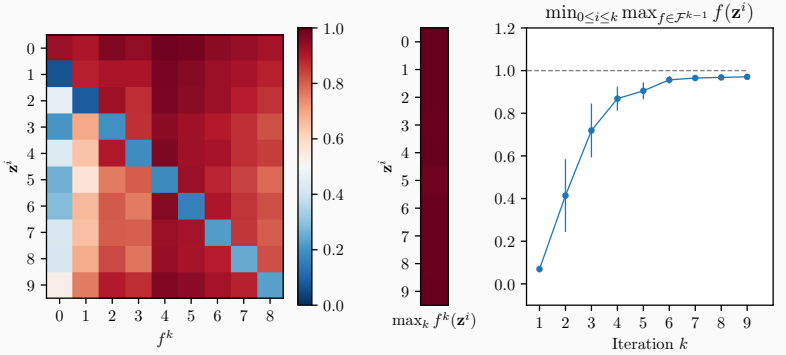
- Heuristic

$$\rho_i^{+,new} = \begin{cases} \rho_i^+ / \alpha & \text{if } \frac{\partial f}{\partial y_i}(\mathrm{y}) < 0, \\ \rho_i^+ & \text{if } \frac{\partial f}{\partial y_i}(\mathrm{y}) = 0, \\ \rho_i^+ \alpha & \text{if } \frac{\partial f}{\partial y_i}(\mathrm{y}) > 0, \end{cases} \quad \text{with } \alpha = 2 \tag{8}$$

Flaws of ADV-EMB

- At some point, it fails at solving

$$z^* = \arg\min_{z \in \mathcal{A}_a} \max_{f \in \mathcal{F}^{k-1}} f(z).$$



$\mathbb{E}[f^k(z_i)]$ $\mathbb{E}[\max_k f^k(z_i)]$ $\mathbb{E}[\arg\min_i \max_k f^k(z_i)]$

Improving the cost map via a back-propagable attack: Backpack

Backpack

Backpack: Back-propagable attack in order to compute $\nabla_\rho \mathbb{E}_y[f(y)]$.

$$x \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow{\text{draw}} \{b_i\}_i \longrightarrow y \longrightarrow \mathbb{E}_y[f(y)]$$
$$\xleftarrow{\nabla_\rho \mathbb{E}_y[f(y)]}$$

Backpack

Backpack: Back-propagable attack in order to compute $\nabla_\rho \mathbb{E}_y[f(\mathrm{y})]$.

$$\mathrm{x} \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow{\text{draw}} \{b_i\}_i \longrightarrow \mathrm{y} \longrightarrow \mathbb{E}_y[f(\mathrm{y})]$$
$$\xleftarrow{\nabla_\rho \mathbb{E}_y[f(\mathrm{y})]}$$

$$\mathrm{x} \longrightarrow \{\rho_i\}_i \xrightarrow[2.\ \frac{\partial \pi}{\partial \rho}]{} \{\pi_i^b\}_i \xrightarrow[1.\ \frac{\partial b}{\partial \pi}]{\text{draw}} \{b_i\}_i \xrightarrow[\frac{\partial y}{\partial b}]{} \mathrm{y} \xrightarrow[\nabla_y \mathbb{E}_y[f(\mathrm{y})]]{} \mathbb{E}_y[f(\mathrm{y})]$$

$$x \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow{\text{draw}} \{b_i\}_i \longrightarrow y \longrightarrow \mathbb{E}_y[f(y)]$$

$$\{\pi_i^b\}_i \xleftarrow{\frac{\partial b_j}{\partial \pi_i^b}} \{b_i\}_i$$

$$x \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow{\text{draw}} \{b_i\}_i \longrightarrow y \longrightarrow \mathbb{E}_y[f(y)]$$
$$\xleftarrow{\frac{\partial b_j}{\partial \pi_i^b}}$$

Replacing the discrete distribution by a smooth approximation by B_τ:

$$x \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow{B_\tau(.,u)} \{\tilde{b}_i\}_i \longrightarrow \tilde{y} \longrightarrow \mathbb{E}[f(\tilde{y})]$$

$$x \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow[\frac{\partial b_i}{\partial \pi_i^b}]{\text{draw}} \{b_i\}_i \longrightarrow y \longrightarrow \mathbb{E}_y[f(y)]$$

Replacing the discrete distribution by a smooth approximation by B_τ:

$$x \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \xrightarrow[\frac{\partial \tilde{b}_i}{\partial \pi_i^b}]{B_\tau(.,u)} \{\tilde{b}_i\}_i \longrightarrow \tilde{y} \longrightarrow \mathbb{E}[f(\tilde{y})]$$

Definition (Hardmax Gumbel Function)

Hardmax Gumbel[13] is a function of $\pi = \{\pi^{-1}, \pi^0, \pi^{+1}\}$ and $\mathrm{g} = \{g^{-1}, g^0, g^{+1}\}$ and simulates, when g is drawn from $G(0,1)$ a drawing from a categorical distribution in $\{-1, 0, 1\}$ with probability distribution π:

$$b = \mathrm{HG}(\boldsymbol{\pi}, \mathrm{g}) = \underset{j \in \{-1,0,+1\}}{\arg\max} \ (g^j + \log \pi^j). \tag{9}$$

[13] Jang, Gu, and Poole 2016.

The softmax function, an approximation of arg max function:

$$\operatorname{softmax}(z_1, \ldots, z_n) = \frac{1}{\sum_{k=1}^{n} e^{z_k}} (e^{z_1}, \ldots, e^{z_n}), \tag{10}$$

because:

$$\lim_{\tau \to 0} \operatorname{softmax}\left(\frac{z_1}{\tau}, \ldots, \frac{z_n}{\tau}\right) = \underbrace{(0, 0 \ldots, 0, 1, 0, \ldots, 0)}_{1 \text{ at position } \arg\max \mathrm{HG}}, \tag{11}$$

τ is called the *temperature*.

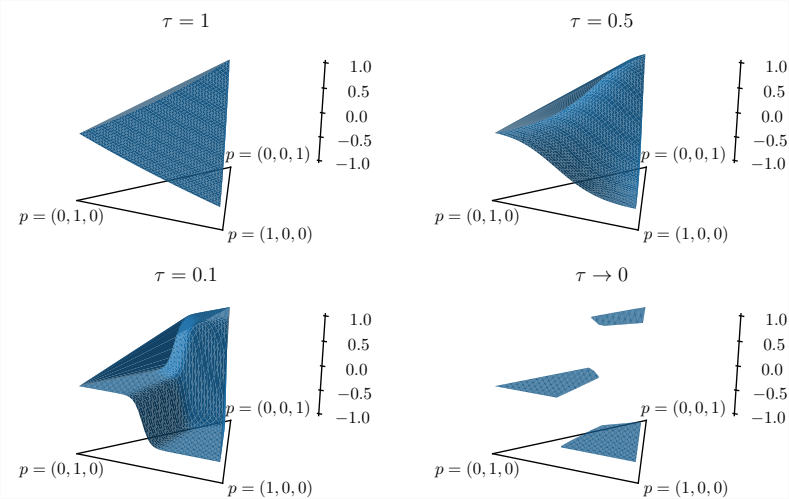
Definition (Softmax Gumbel Function)

Hardmax Gumbel function can be approximated by the following Softmax Gumbel function:

$$\tilde{b}_\tau = \mathrm{SG}_\tau(\boldsymbol{\pi}, \mathrm{g}) = \sum_{j \in \{-1,0,+1\}} j\, z^j, \tag{12}$$

$$\text{with } \mathrm{z} = \mathrm{softmax}\left(\frac{\mathrm{g} + \log \boldsymbol{\pi}}{\tau}\right). \tag{13}$$

1. Approximation of discrete modifications with Softmax Gumbel distribution



2. $\frac{\partial \pi}{\partial \rho}$ computed implicitly

$$\mathrm{x} \longrightarrow \{\rho_i\}_i \underset{\xleftarrow{\frac{\partial \pi}{\partial \rho}}}{\longrightarrow} \{\pi_i^b\}_i \longrightarrow \{b_i\}_i \longrightarrow \mathrm{y} \longrightarrow \mathbb{E}_y[f(\mathrm{y})]$$

2. $\frac{\partial \pi}{\partial \rho}$ computed implicitly

$$\mathrm{x} \longrightarrow \{\rho_i\}_i \underset{\overleftarrow{\frac{\partial \pi}{\partial \rho}}}{\longrightarrow} \{\pi_i^b\}_i \longrightarrow \{b_i\}_i \longrightarrow \mathrm{y} \longrightarrow \mathbb{E}_y[f(\mathrm{y})]$$

Probabilities are a function of costs ρ and λ

$$\pi_i^j = P_i^j(\rho_i, \lambda) = \frac{e^{-\lambda \rho_i^j}}{\sum_{k \in \{-1,0,1\}} e^{-\lambda \rho_i^k}}, \quad j \in \{-1, 0, 1\} \tag{14}$$

The chain rule gives:

$$\frac{d\boldsymbol{\pi}}{d\boldsymbol{\rho}} = \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\rho}} + \frac{\partial \boldsymbol{\pi}}{\partial \lambda} \nabla_{\boldsymbol{\rho}} \lambda \tag{15}$$

where $\lambda = \Lambda(\rho, |\mathrm{m}|))$ is an (implicit) function of ρ and $|m|$ stemming from the entropy constraint: $H(\boldsymbol{\pi}) = |\mathrm{m}|$

2. $\frac{\partial \pi}{\partial \rho}$ computed implicitly

The total derivative of H:

$$\nabla_{\boldsymbol{\rho}} H(\boldsymbol{\pi}) \frac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\rho}} + \frac{\partial H(\boldsymbol{\pi})}{\partial \lambda} \nabla_{\boldsymbol{\rho}} \lambda = 0 \tag{16}$$

2. $\frac{\partial \pi}{\partial \rho}$ computed implicitly

The total derivative of H:

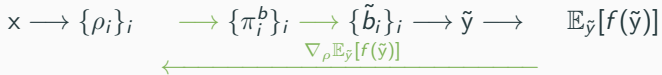
$$\nabla_{\boldsymbol{\rho}} H(\boldsymbol{\pi}) \frac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\rho}} + \frac{\partial H(\boldsymbol{\pi})}{\partial \lambda} \nabla_{\boldsymbol{\rho}} \lambda = 0 \tag{16}$$

Explicit gradient of λ accessible through differentiation of the entropy constraint.

$$\nabla_{\boldsymbol{\rho}} \lambda = - \left(\frac{\partial H(\boldsymbol{\pi})}{\partial \lambda} \right)^{-1} \nabla_{\boldsymbol{\rho}} H(\boldsymbol{\pi}). \tag{17}$$

Overall algorithm of Backpack

Backpack: Back-propagable attack.



Overall algorithm of Backpack

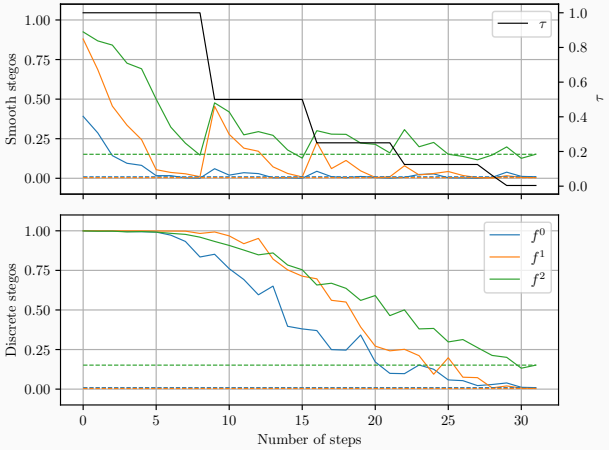
Backpack: Back-propagable attack.

$$x \longrightarrow \{\rho_i\}_i \longrightarrow \{\pi_i^b\}_i \longrightarrow \{\tilde{b}_i\}_i \longrightarrow \tilde{y} \longrightarrow \mathbb{E}_{\tilde{y}}[f(\tilde{y})]$$
$$\longleftarrow \quad \nabla_\rho \mathbb{E}_{\tilde{y}}[f(\tilde{y})]$$

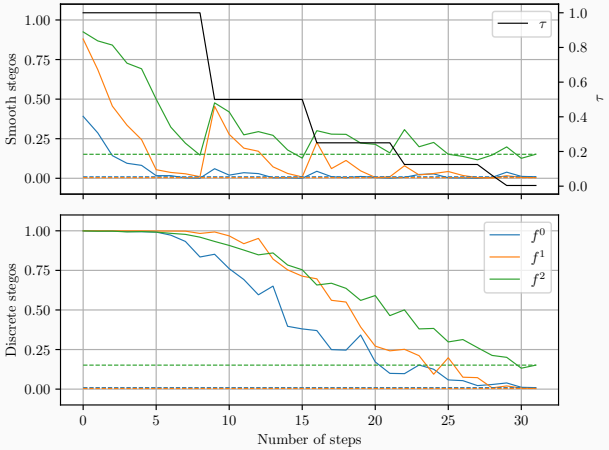
Iterative gradient descent, while playing with the temperature τ, to optimize the smooth output w.r.t. best detector:

$$\nabla_\rho \mathbb{E}_{\tilde{y}}[f(\tilde{y})], \quad f = \arg\max_{f \in \mathcal{F}^{k-1}} \mathbb{E}_{\tilde{y}}[f(\tilde{y})] \tag{18}$$

Difference between smooth and discrete stegos



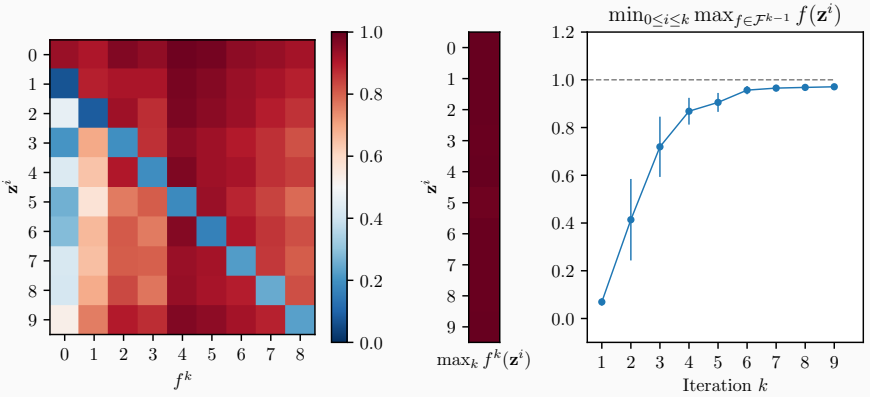
Difference between smooth and discrete stegos



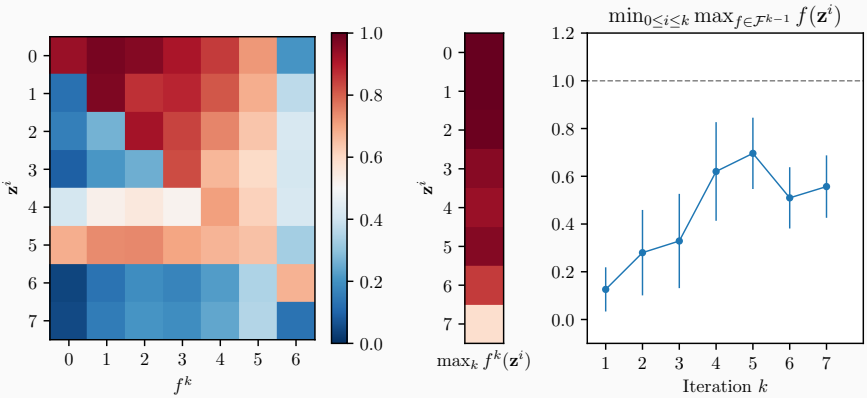
Hyper-parameters:

- Number of samples
- Temperature policy
- Number of steps for the gradient descent
- Stopping condition (here depends on the value of the detectability of covers)

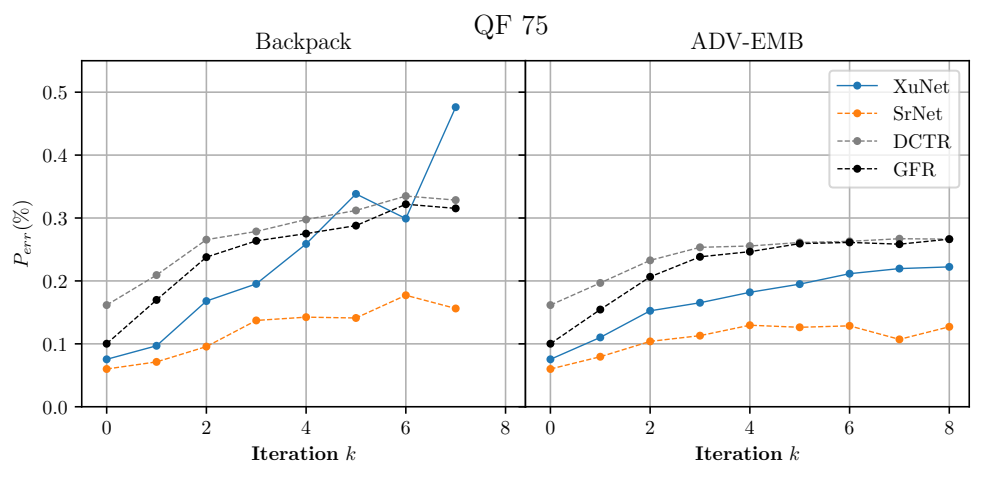
Evolution of min max protocol with ADV-EMB



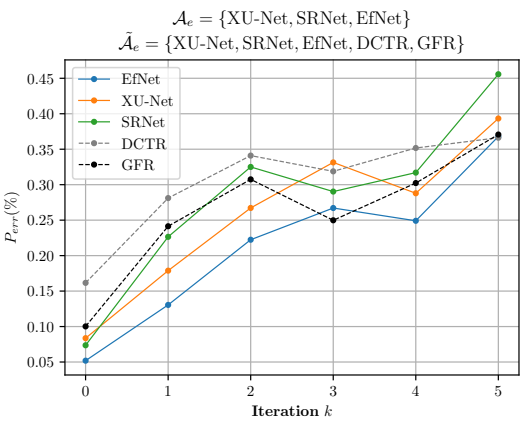
Evolution of min max protocol with Backpack



Results - Comparison ADV-EMB and Backpack for JPEG at QF 75 and at payload 0.4 bpnzAC



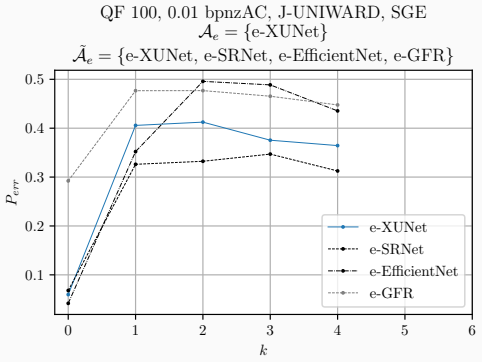
Results - With three targeted models, for images at QF 75 and payload 0.4 bpnzAC



EfNet	**XU-Net**	**SRNet**	DCTR	GFR
+31.6 pts	+31.0 pts	+38.2 pts	+20.5 pts	+27.0 pts

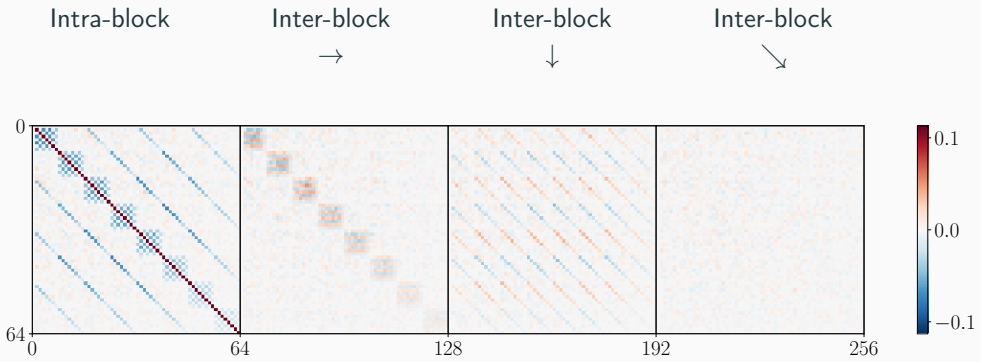
Results - Protocol with Backpack, at QF 100, with e-XUNet

Classifier using the rounding error of decompression[14]



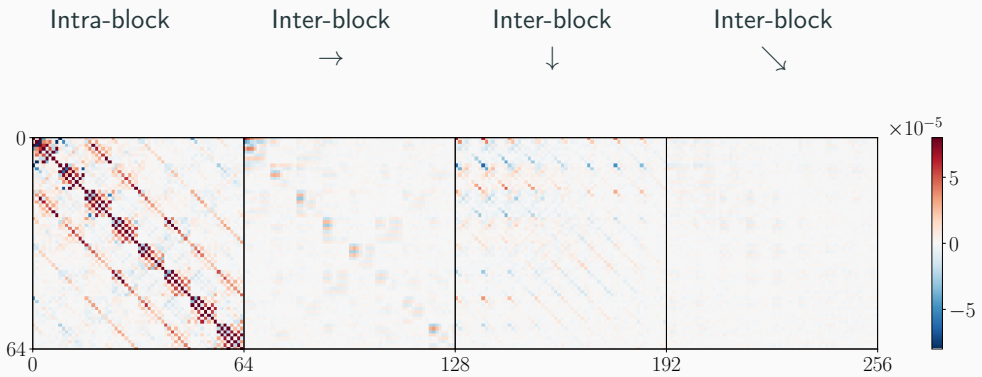
[14] Jan Butora and Jessica Fridrich (2019). "Reverse JPEG compatibility attack". In: *IEEE Transactions on Information Forensics and Security* 15, pp. 1444–1454.

Interpretation - Natural correlations between DCT modes[15]

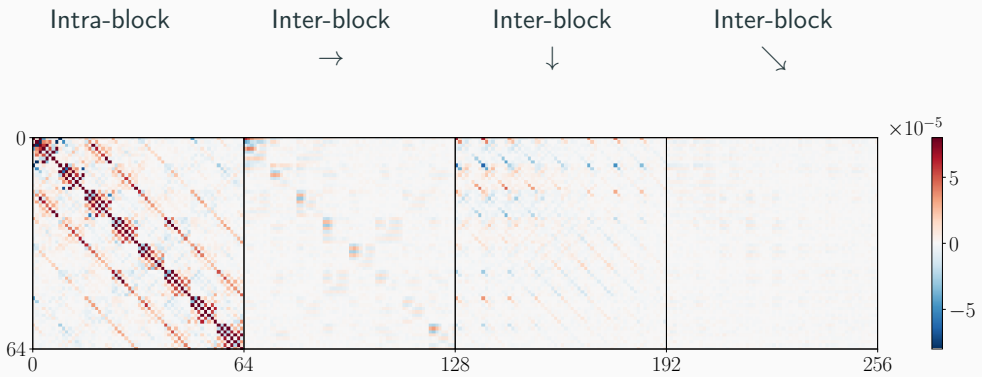


[15] Théo Taburet et al. (2019). "Computing dependencies between DCT coefficients for natural steganography in JPEG domain". In: *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, pp. 57–62.

Interpretation - Observed correlations between changes of DCT modes



Interpretation - Observed correlations between changes of DCT modes



For symmetric costs, correlations between DCT modes are equal to 0.

Conclusion and perspectives

Conclusions

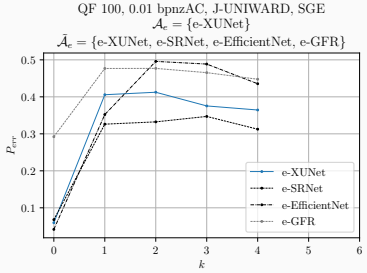
Thanks to adopting recent methods from

- Game theory
- and machine learning

we have greatly improved the accuracy of steganography.

Perspectives

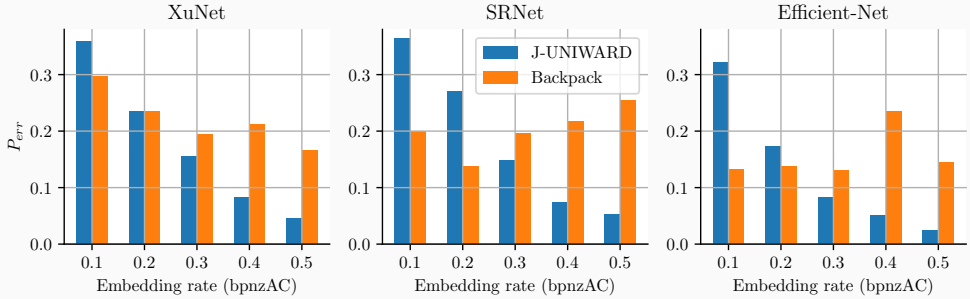
- Backpack very computationnaly costly, and gets worst with increasing iterations of the protocol.



  - Distillation: to merge the knowledge of multiple classifiers
  - Hyper-parameters to optimize for Backpack: number maximum of steps, learning rate, stopping condition.

Perspectives

Experiment at payload 0.4 bpnzAC with targeted XUNet, SRNet, and Efficient-Net.



- Payload dependent: specific to the payload
- Source dependent: the whole protocol should be re-run for every cover source