ZASTOSOWANIE SZTUCZNYCH SIECI NEURONOWYCH W KRYPTOLOGII

ADAM ŻYCHOWSKI

Kryptologia - dziedzina wiedzy o przekazywaniu informacji w sposób zabezpieczony przed niepowołanym dostępem

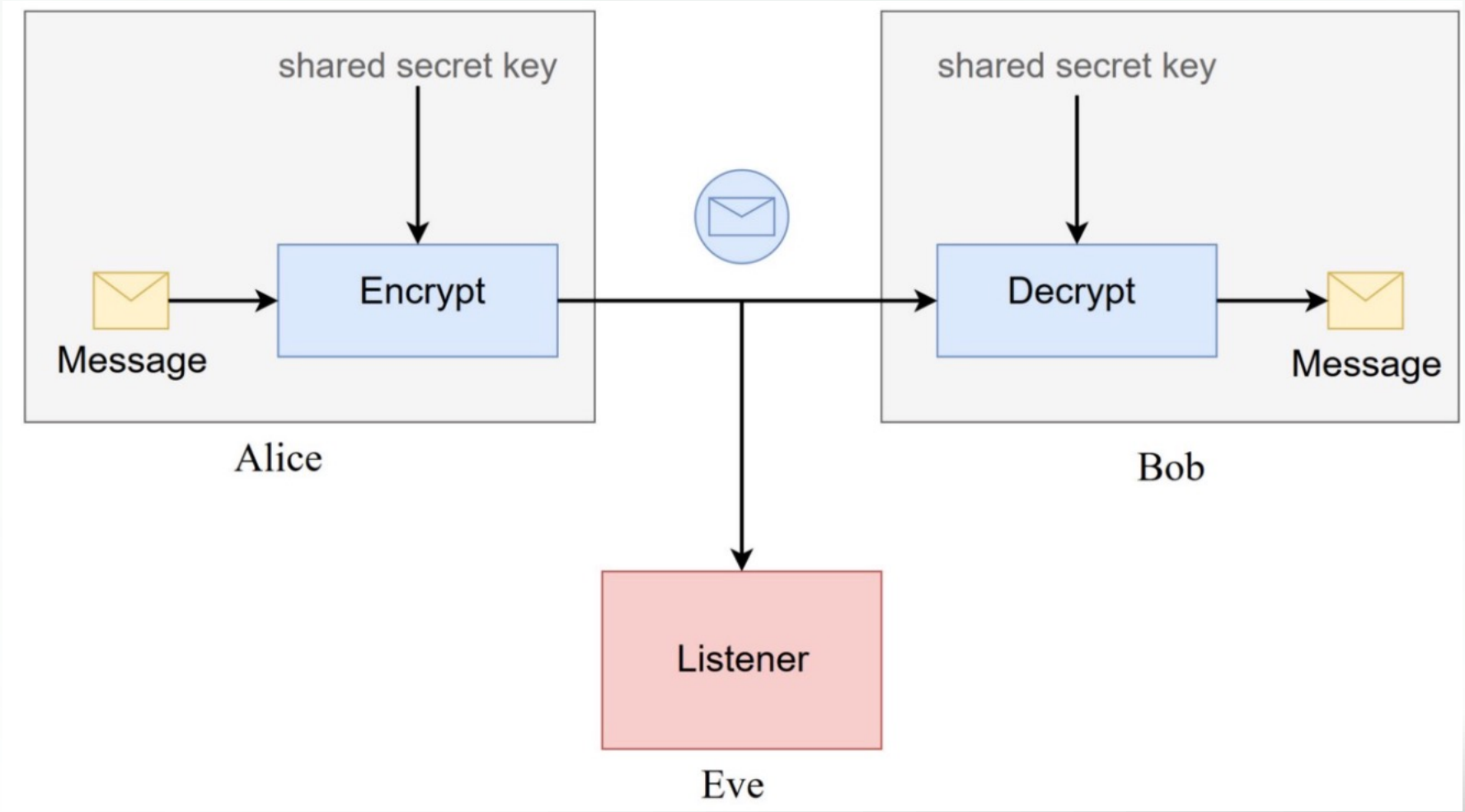
- Szyfrowanie wiadomości (kryptografia)
- Deszyfrowanie wiadomości (kryptoanaliza)
- Nauka sieci neuronowych na zaszyfrowanych danych

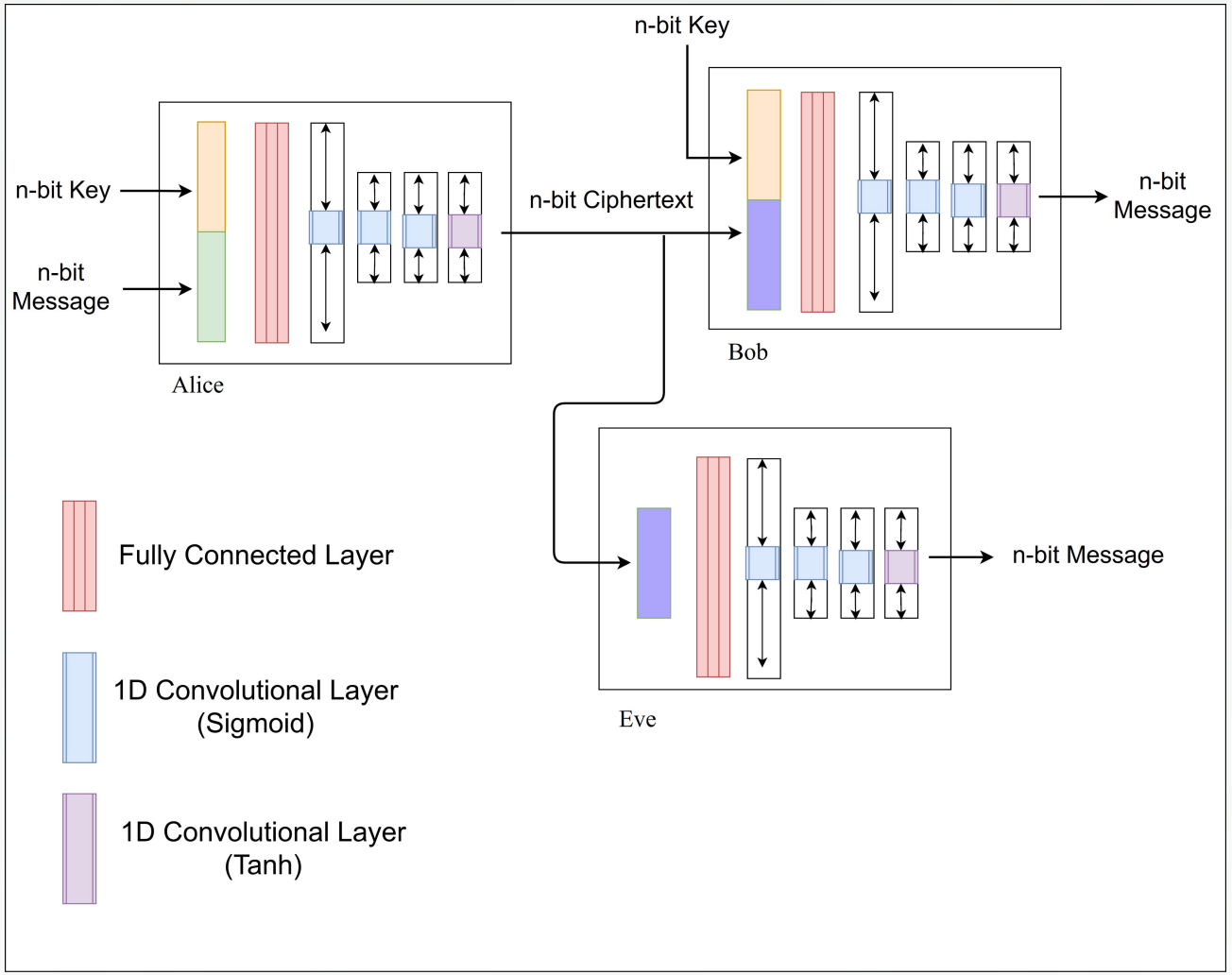
KRYPTOGRAFIA

przekazywanie wiadomości w sposób zabezpieczony (niejawny) przed niepowołanym dostępem, sposoby utajniania informacji

dziedzina bardzo stara (np. Szyfr Cezara I wiek p.n.e.), ale nadal o wielkim znaczeniu (hasła komputerowe, transakcje, protokoły komunikacji internetowej, wojskowość)

SZYFROWANIE SYMETRYCZNE

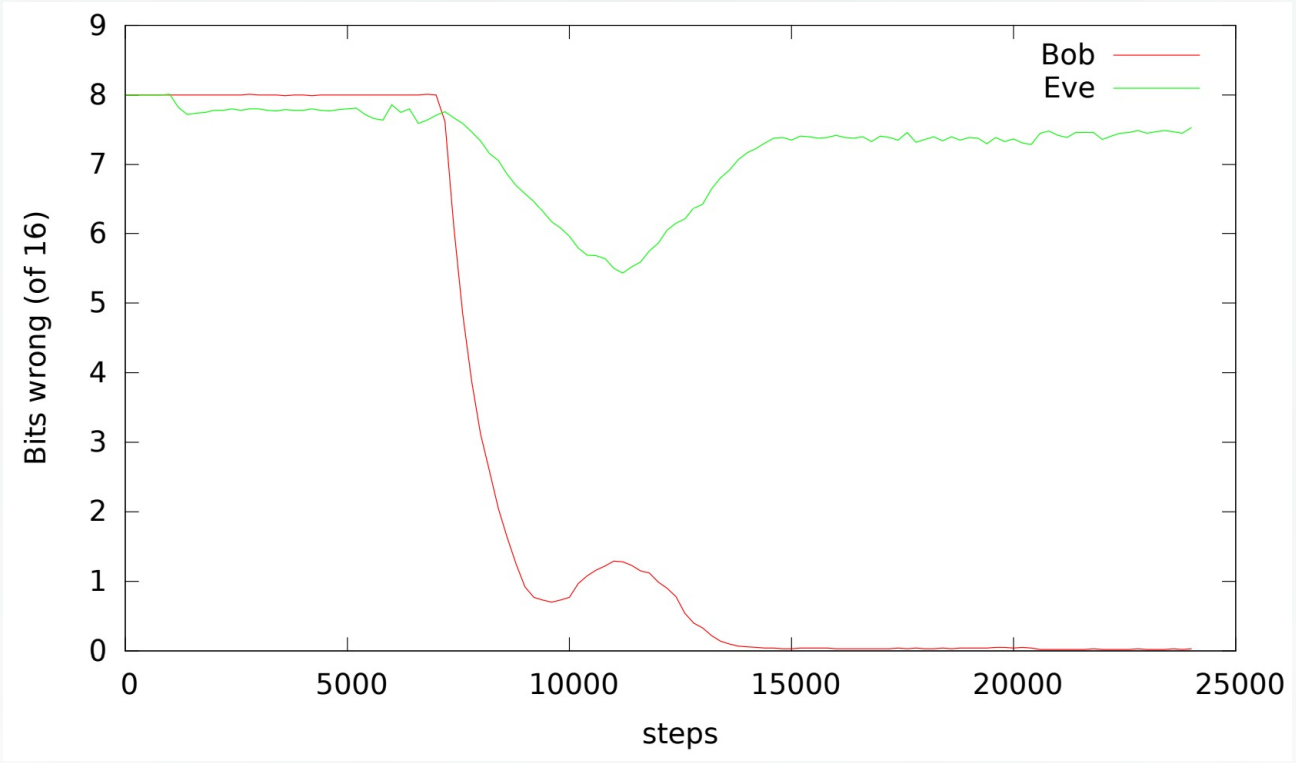




$$Eve\_error = L_1(eve\_out - message)$$

$$Bob\&Alice\_error = L_1(bob\_out - message) + (\frac{N}{2} - Eve\_error)^2$$

Abadi et al. 2016

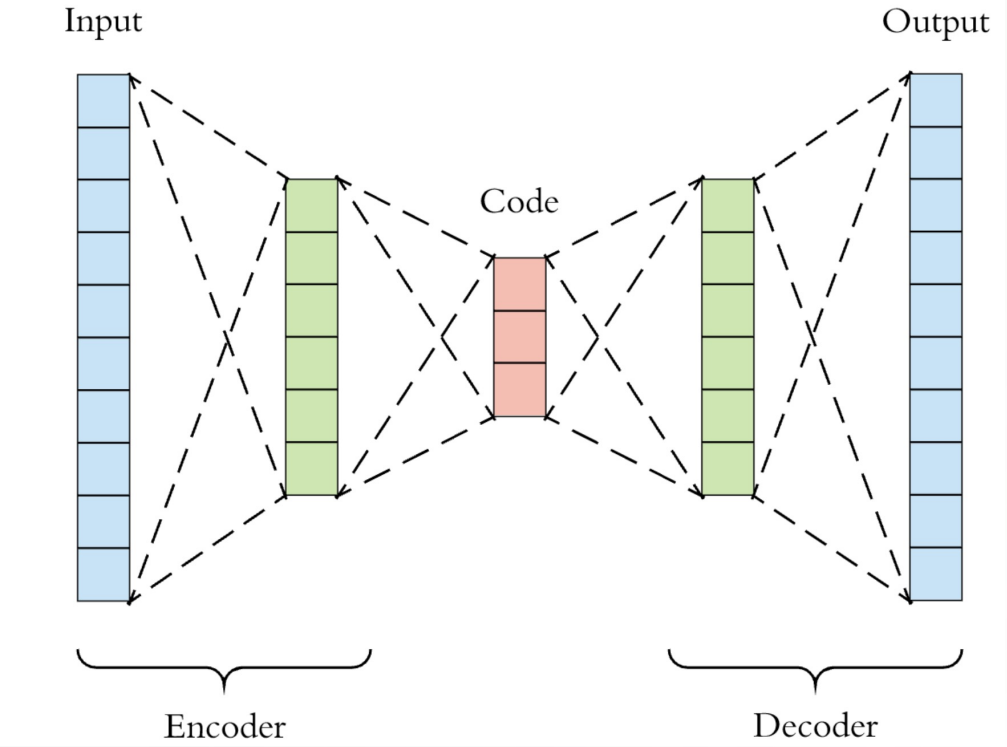


"We do not prescribe specific cryptographic algorithms to these neural networks; instead, we train end-to-end, adversarially. We demonstrate that the neural networks can learn how to perform forms of encryption and decryption"

Abadi et al. 2016

AUTOENKODER

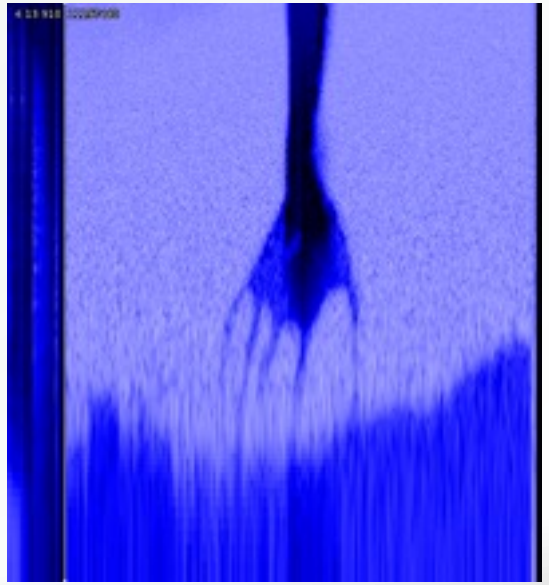
- użycie środkowej warstwy jako zakodowana wiadomość
- sekret: parametry modelu, zbiór treningowy, sekretny element dodawany w treningu



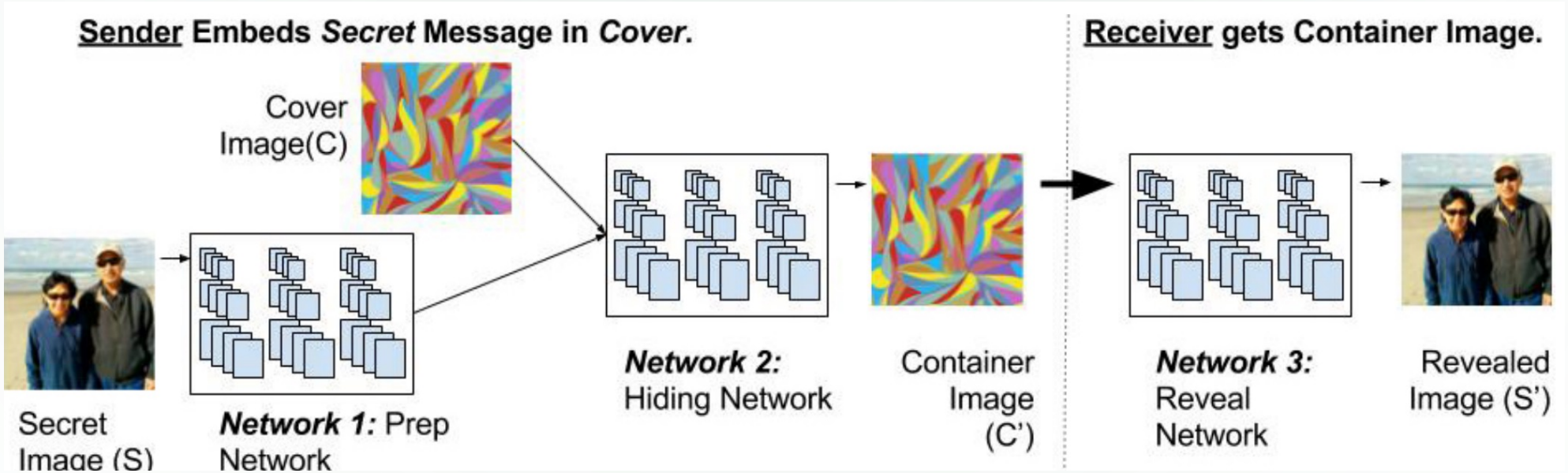
Sagar et al. 2019

STEGANOGRAFIA

Ukrywanie wiadomości w taki sposób, aby obecność komunikatu nie została wykryta.



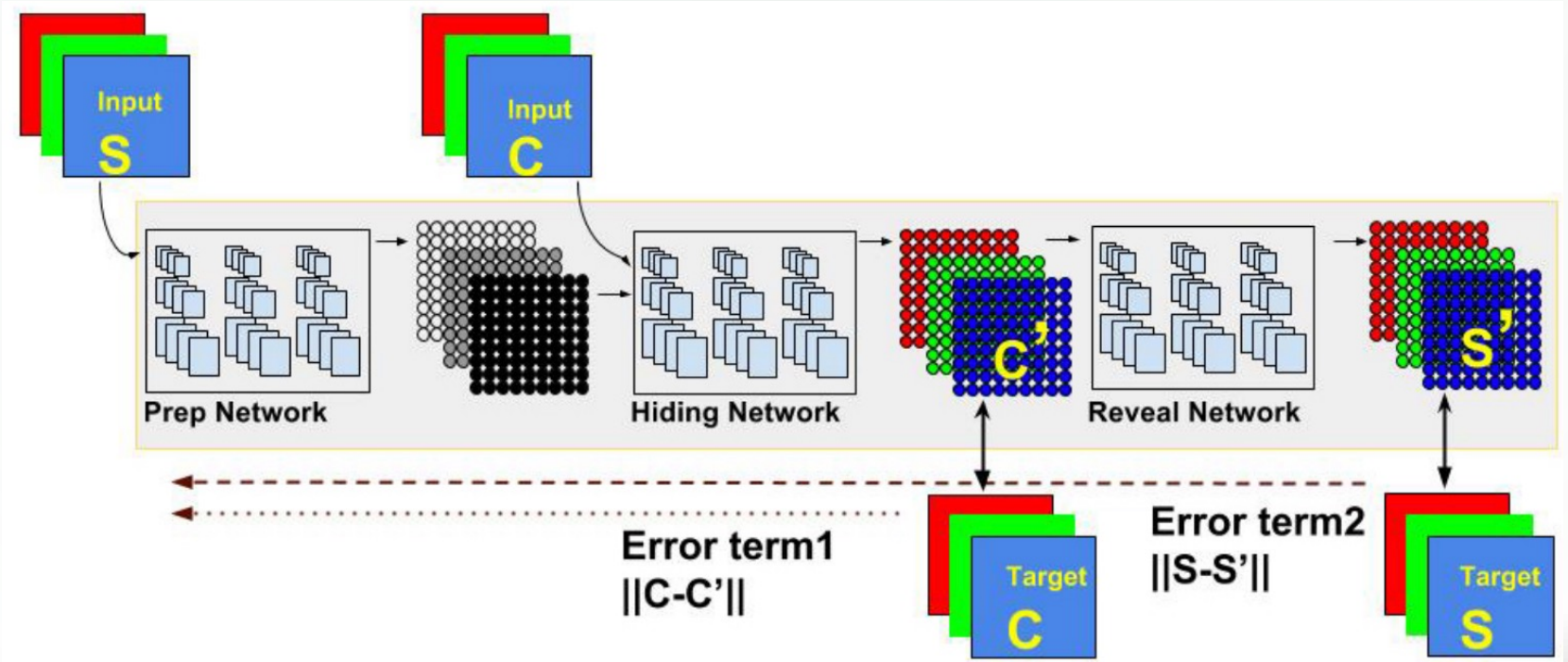
DEEP STEGANOGRAPHY

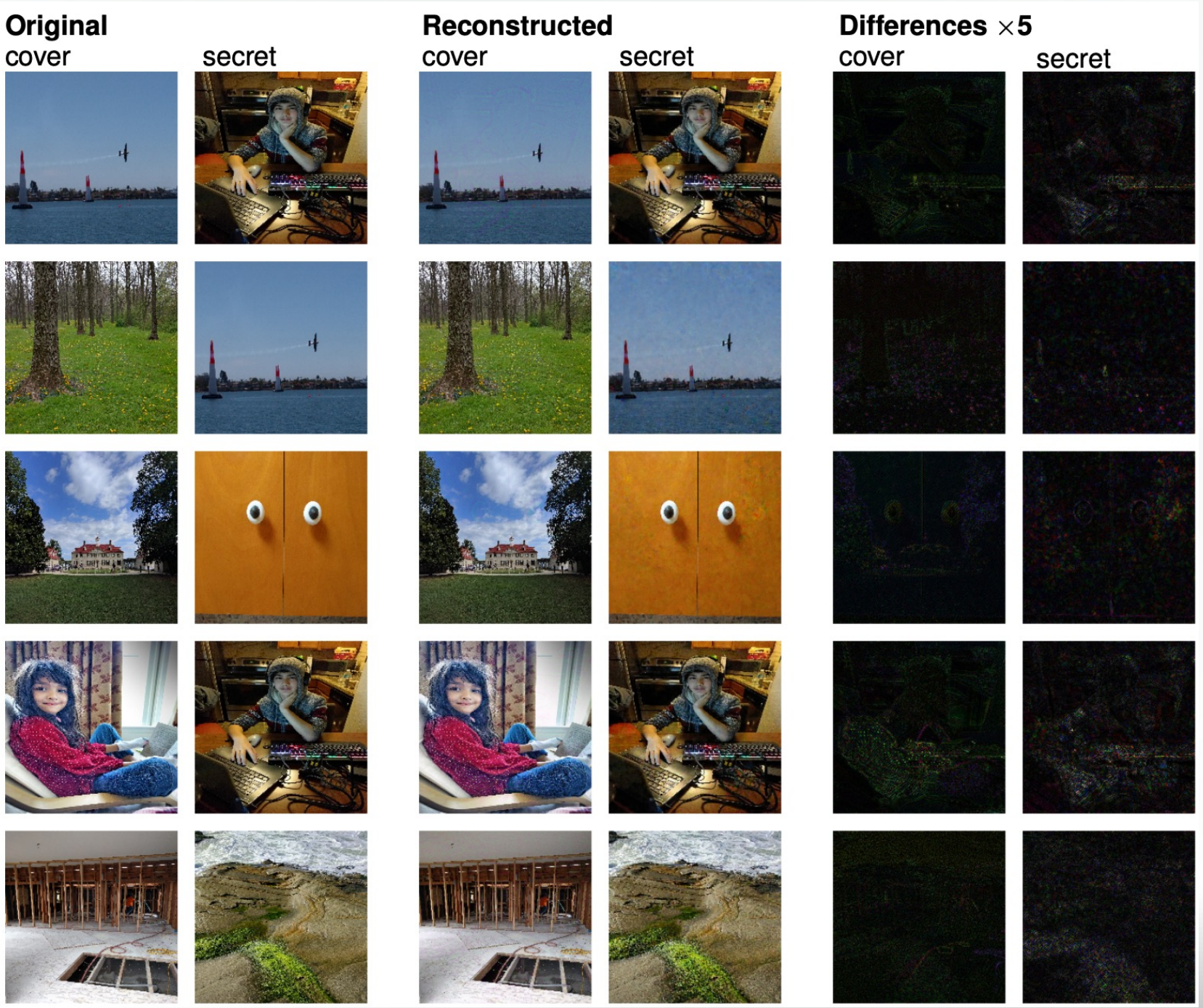


Baluja 2017

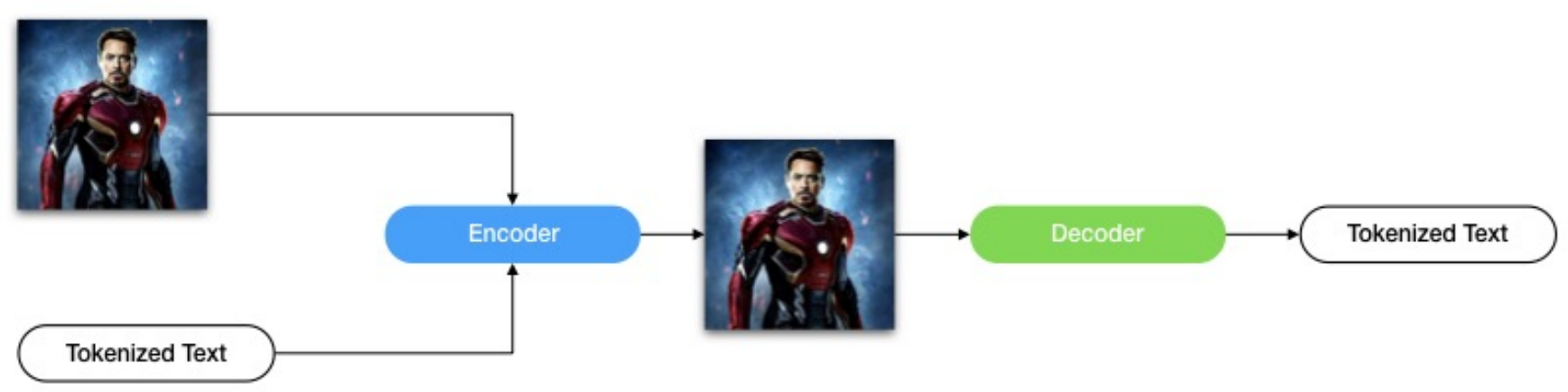
DEEP STEGANOGRAPHY

Error = $||c-c'|| + \beta||s-s'||$

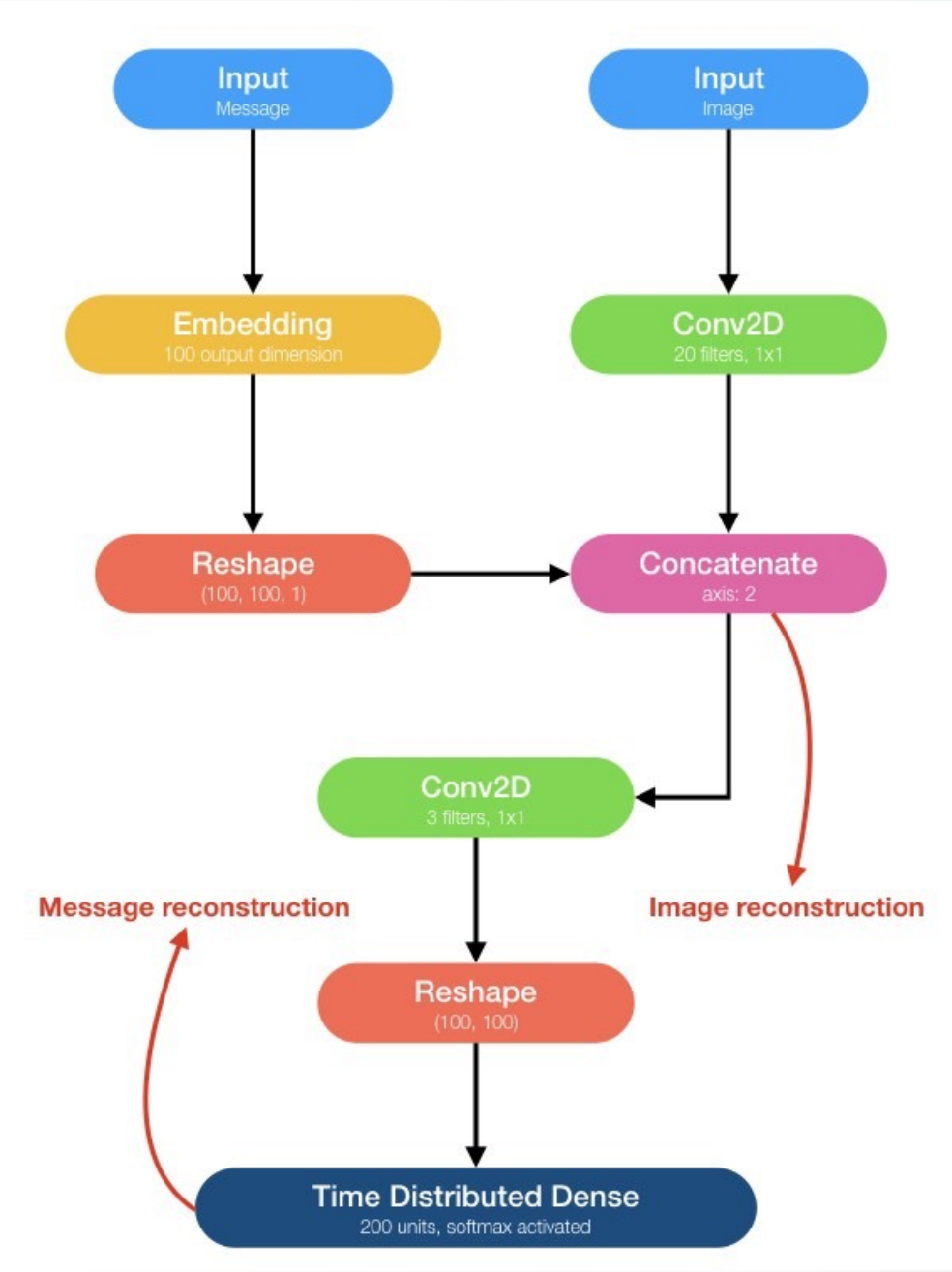



Original
cover
secret
Reconstructed
cover
secret
Differences ×5
cover
secret

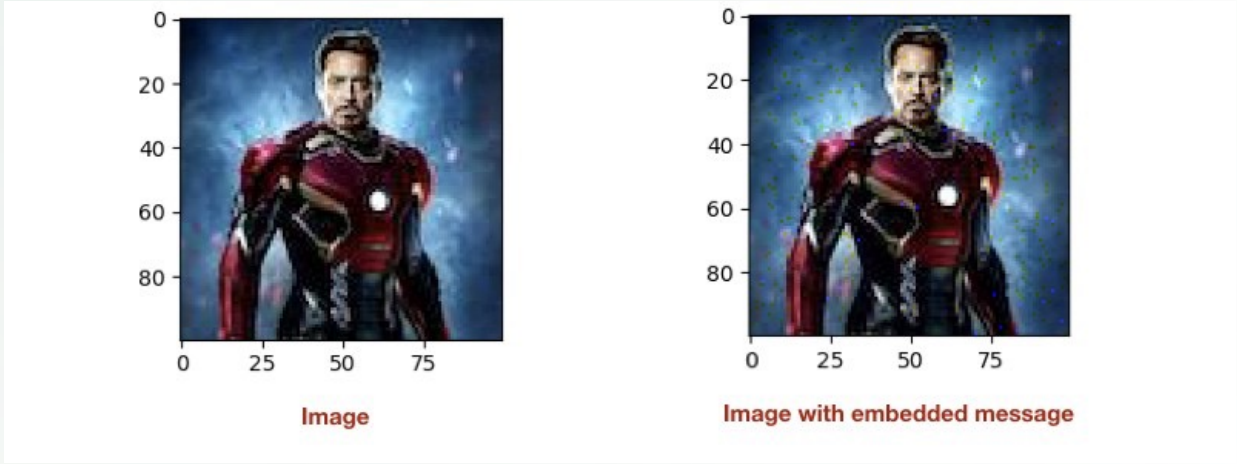
TEXT IMAGE STEGANOGRAPHY



https://towardsdatascience.com/neural-cryptography-7733f18184f3



https://towardsdatascience.com/neural-cryptography-7733f18184f3



Anthony Edward Tony Stark is a character portrayed by Robert Downey Jr. in the MCU film franchise

KRYPTOANALIZA

- działania, które mają na celu uzyskanie wiadomości wrażliwej (tajnej), w tym:
  łamanie szyfrów

f(tekst_jawny, klucz) = tekst_zaszyfrowany

SZYFR VIGENÈRE

CALCUL

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

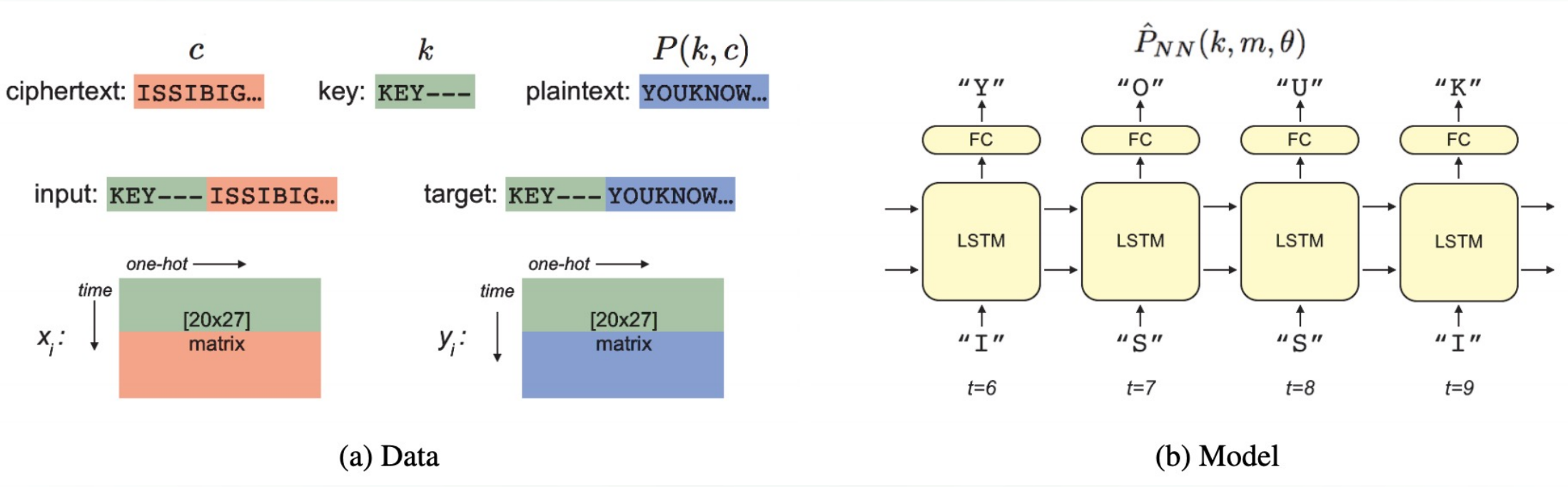
AUTOKLUCZ (AUTOKEY)

- szyfrowany tekst służy jako klucz
- rozwiązuje problem skończonej (ustalonej) długości klucza

Przykład:

```
Plaintext:  ATTACK AT DAWN...
Key:        QUEENL YA TTACK AT DAWN....
Ciphertext: QNXEPV YT WTWP...
```

SIECI REKURENCYJNE



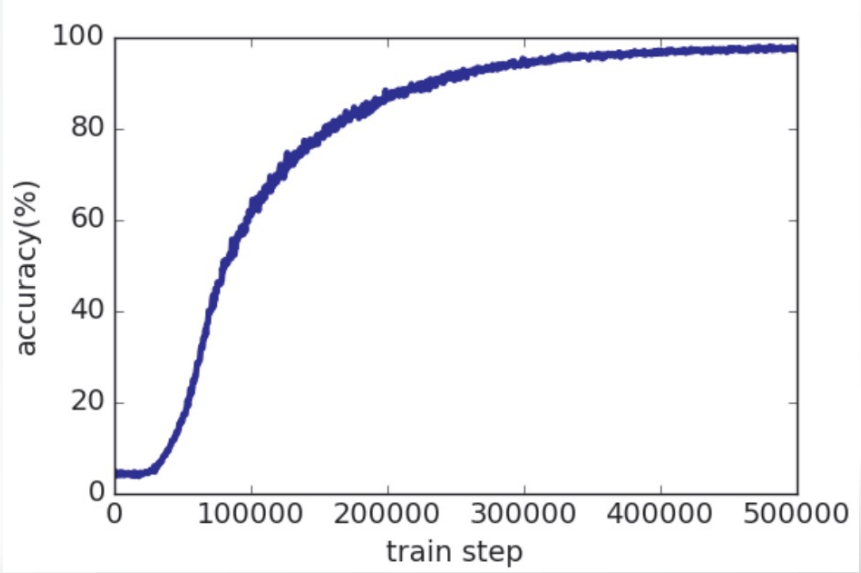
Greydanus 2017

NAUKA I WYNIKI

- 10^7 danych uczących
- losowo wygenerowane teksty o długości 20 znaków

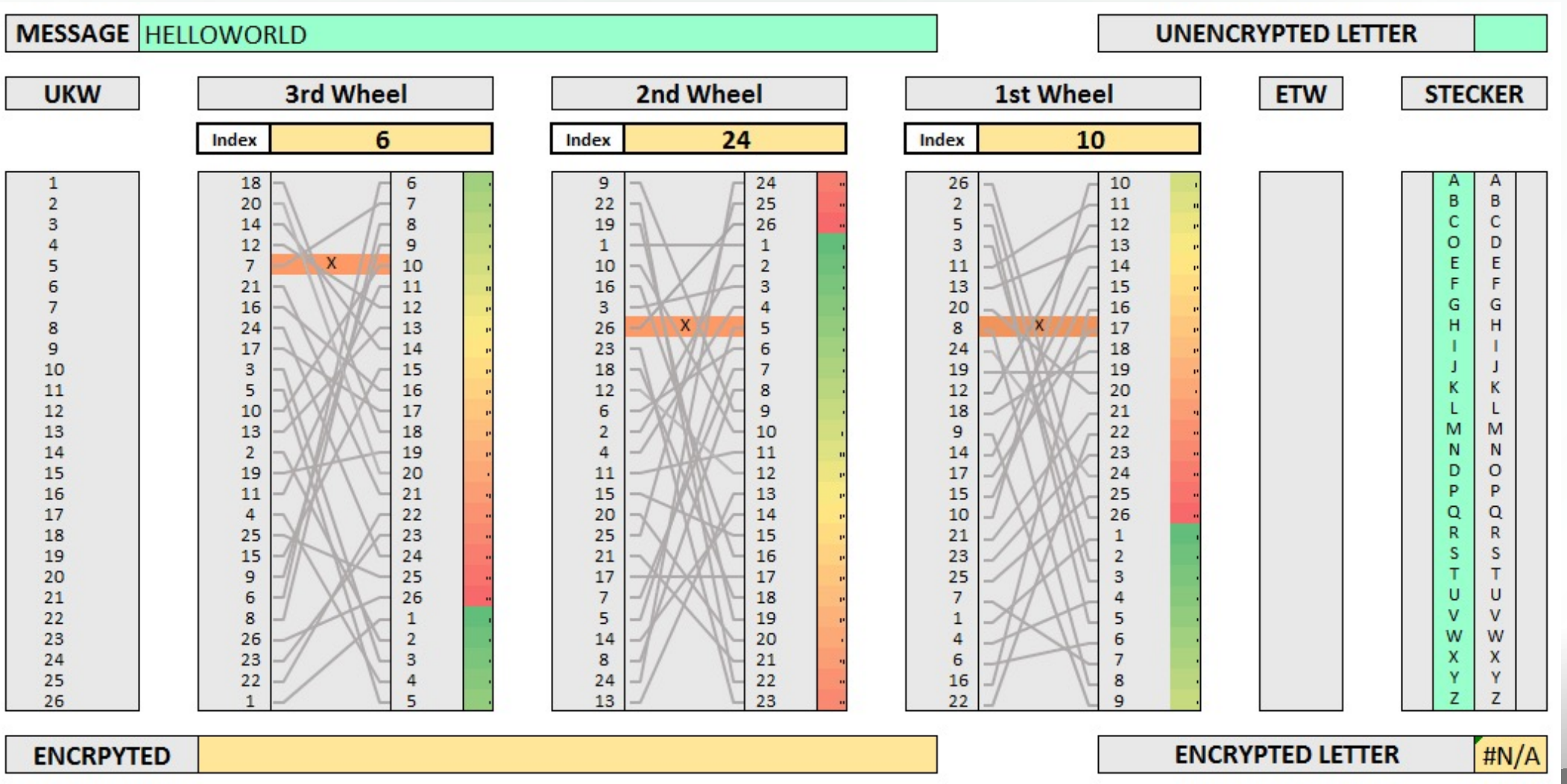
- wszystkich możliwości: 26^20 ~= 10^28

```
[vigenere-nn]>>>python main.py --train False
===== COUNTING MODEL PARAMETERS =====
Model overview:
        variable "model/W_fc1:0" has 5400 parameters
        variable "model/model_cell0/LSTMCell/W_0:0" has 181600 parameters
        variable "model/model_cell0/LSTMCell/B:0" has 800 parameters
Total of 187800 parameters
=====================================
loaded model: models/model.ckpt-150000
accuracy is: 99.770833%
plaintext is:  'WYGTY-YOUKNOWNOTHINGJONSNOW'
ciphertext is: 'WYGTY-UMADLKUTHRDGTZHKLYGMS'
prediction is: 'WYGTY-YOUKNOWNOTHINGJONSNOW'
```



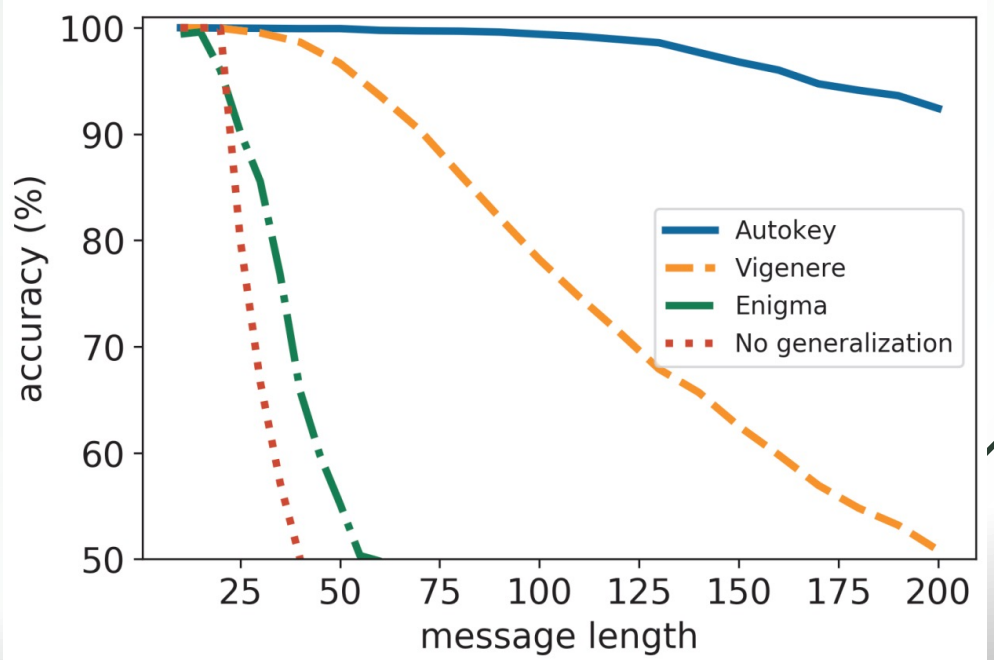
ENIGMA

"Learning the enigma with recurrent neural networks" Greydanus 2017



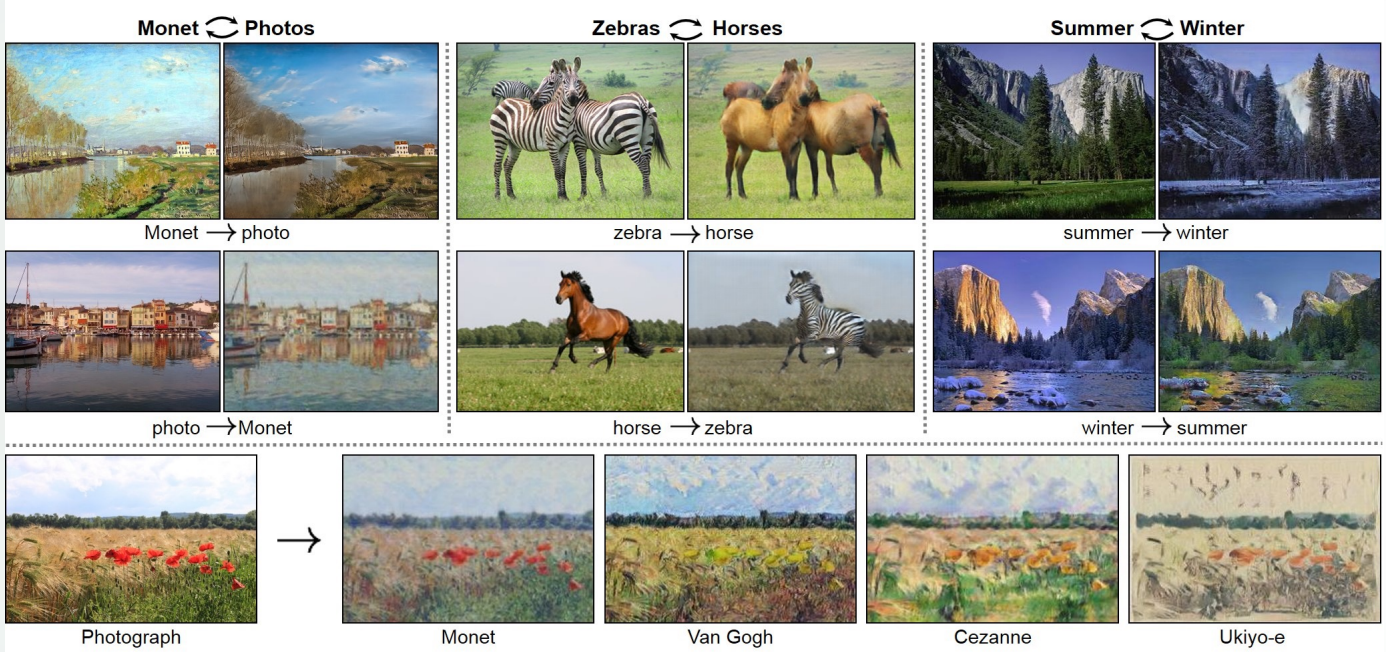
ENIGMA WYNIKI

```
[enigma-rnn]>>>python main.py --train False
===== COUNTING MODEL PARAMETERS =====
Model overview:
        variable "model/W_fc1:0" has 78000 parameters
        variable "model/model_cell0/LSTMCell/W_0:0" has 36312000 parameters
        variable "model/model_cell0/LSTMCell/B:0" has 12000 parameters
Total of 36402000 parameters
=====================================
loaded model: models/model.ckpt-1042500
accuracy is: 96.825000%
plaintext is:  '---YOUKNOWNOTHINGJONSNOW'
ciphertext is: 'EKWJIVMGGJRQXUGBOXQVZXHE'
prediction is: 'EKWYOUWNOWNOTHINGJONSNOW'
```

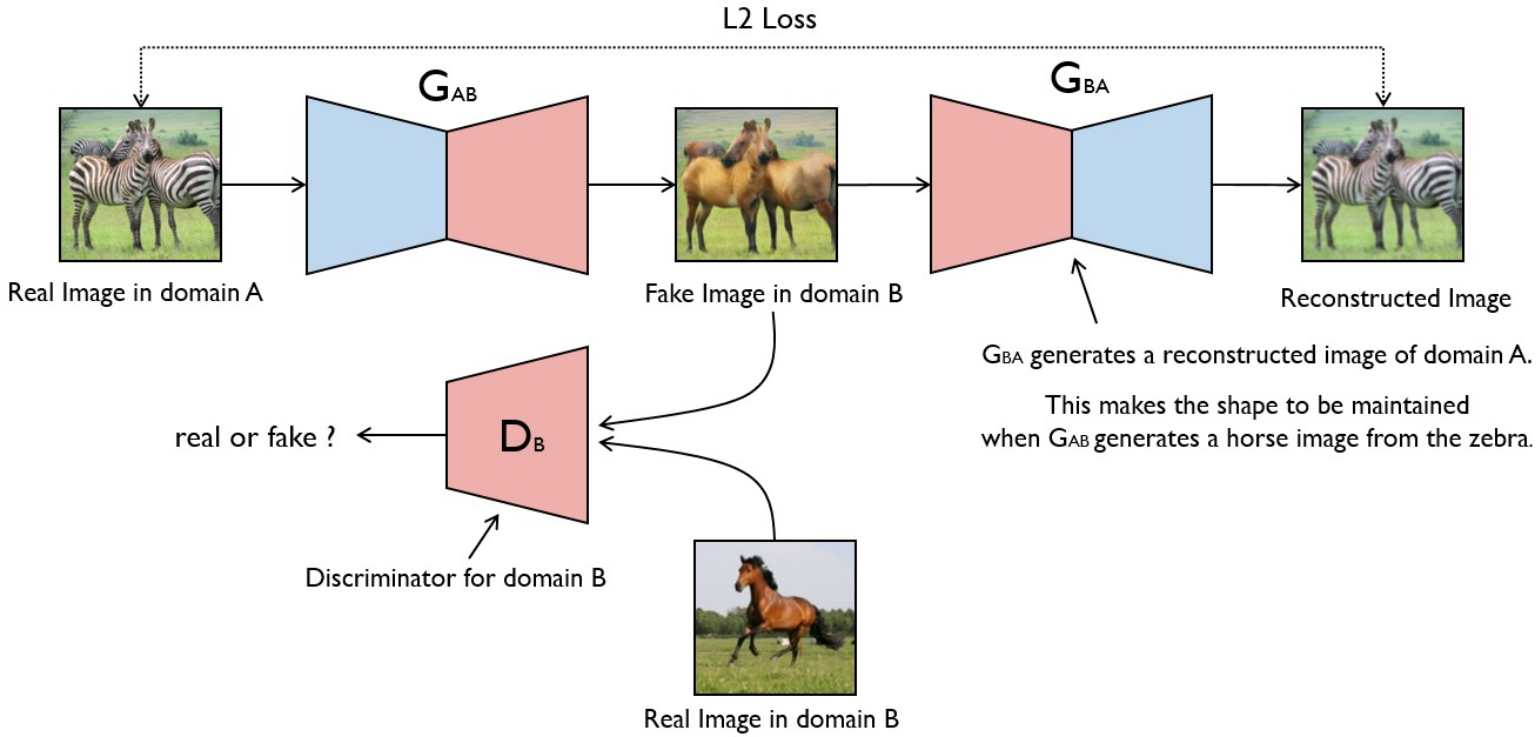


CYCLE GAN

- brak potrzeby par treningowych
- Nauka: 57340 zdań, w sumie ponad milion słów



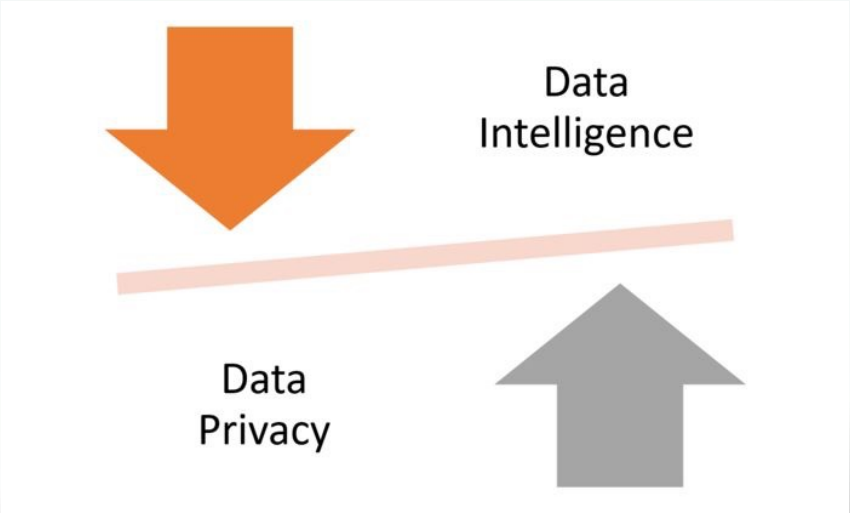
- Shift: ~ 99%
- Vigenère: 80-99%



Gomez et al. 2018

NAUKA NA ZASZYFROWANYCH DANYCH

- istotna i szybko zyskująca na popularności dziedzina badań
- szerokie zastosowania praktycznie: dane medyczne, bankowe, prywatność, tajemnice korporacyjne
- przenoszenie obliczeń związanych z inteligencją obliczeniową do chmury (zewnętrznych dostawców urządzeń)
- anonimizacja to za mało



Cel:

- możliwość zaszyfrowania danych lokalnie i przesłanie ich ”na zewnątrz”
- użyteczność zaszyfrowanych danych (np. nauka sieci neuronowej)
- możliwość odczytania wyniku jedynie po stronie właściciela danych

PEŁNE SZYFROWANIE HOMOMORFICZNE

- H(a+b) = H(a)+H(b)
- H(a*b) = H(a)*H(b)
- H(a*const) = H(a)*const

Gentry, 2009

PEŁNE SZYFROWANIE HOMOMORFICZNE

- H(a+b) = H(a)+H(b)
- H(a*b) = H(a)*H(b)
- H(a*const) = H(a)*const

Wielkość danych: 1 bit informacji zajmuje około 1MB

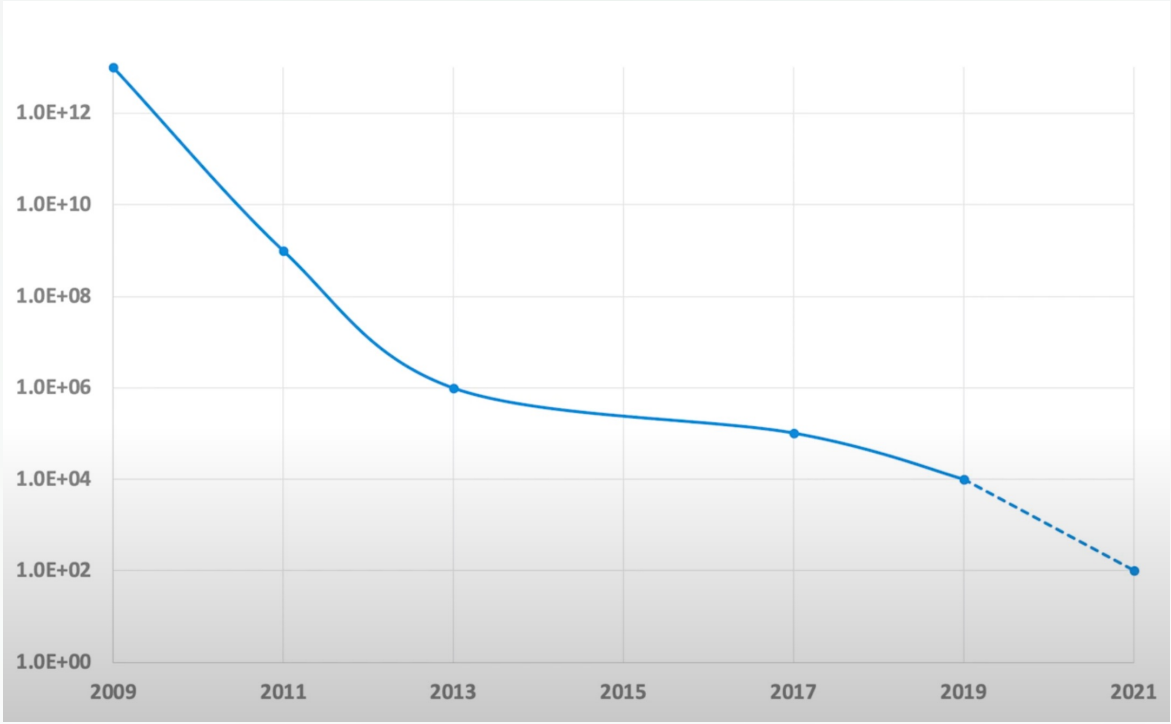
Czas obliczeń: pomnożenie dwóch 32-bitowych liczb zajmie około 10 minut

Gentry, 2009

SZYFROWANIE HOMOMORFICZNE - HISTORIA

- 2009: uznane za niepraktyczne
- 2011: Microsoft poprawia własności szyfrowania, wprowadza “praktyczne szyfrowanie homomorficzne” znacznie przyspieszające obliczenia
- 2016: CryptoNets – pierwsze sieci neuronowe wykorzystujące szyfrowanie homomorficzne
- 2018 - bilbioteka Microsoft SEAL, która popularyzuje i znacznie ułatwia używanie szyfrowania homomorficznego (*sealcrypto.org*)

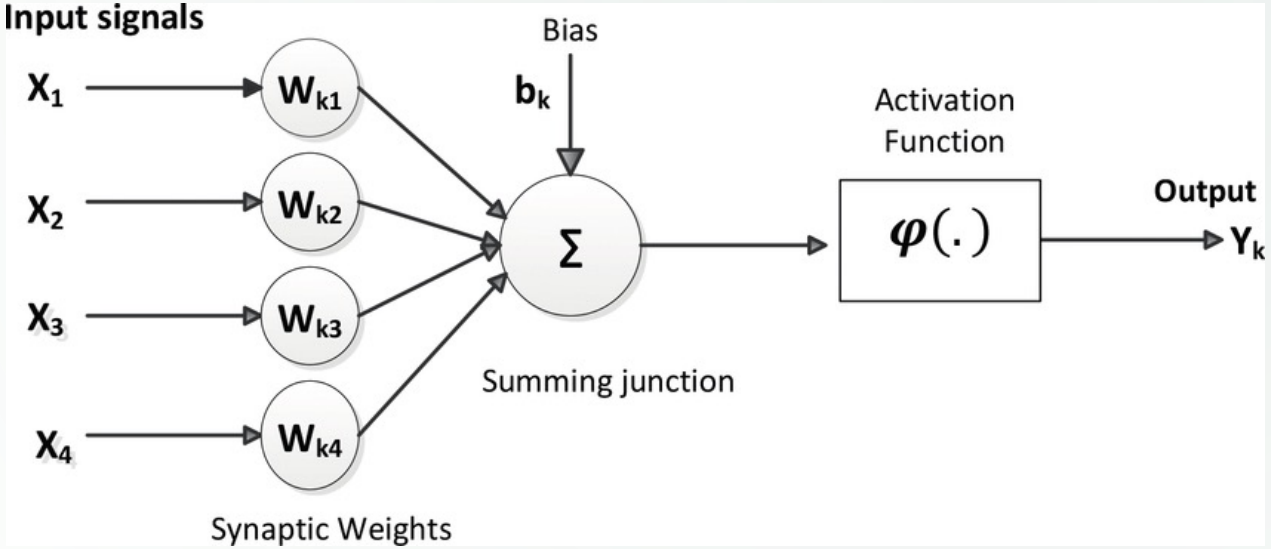
SZYBKOŚĆ SZYFROWANIA HOMOMORFICZNEGO



skala logarytmiczna!

Kristin Lauter,Private AI: Machine Learning on Encrypted Data, 2020

CRYPTO NETS



Mnożenia wag i sumowanie → ok
Problem z funkcją aktywacji – nieliniowa, wielomianowa → f. kwadratowa

dodatkowe optymalizacje: zamiana danych wejściowych na wielomiany – działania na wielomianach znacznie przyspieszają obliczenia, możliwość obliczeń równoległych (np. wielomiany stopnia 8192)

CRYPTO NETS

MNIST: 99% accuracy

Table 2: The performance of CryptoNet for MNIST

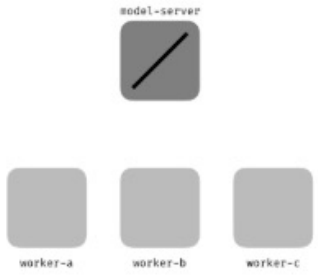
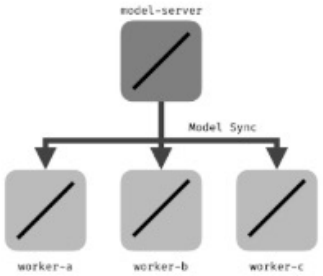
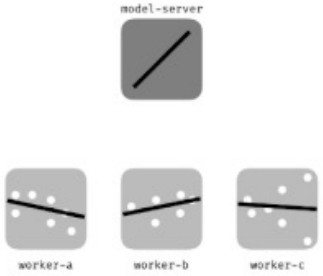
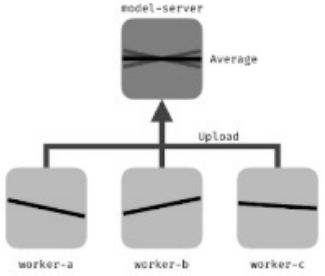
Stage	Latency	Additional latency per each instance in a batch
Encoding+Encryption	122 seconds	0.060 seconds
Network application	570 seconds	0
Decryption+Decoding	5 seconds	0.046 seconds

	Message size	Size per instance
Owner → Cloud	588 MB	73.5 KB
Cloud → Owner	7.5 MB	0.94 KB

Gilad-Bachrach et al. 2016

FEDERATED LEARNING

Modele douczane są lokalnie na urządzeniach końcowych – wszystkie dane pozostają u użytkowników, przesyłane są tylko modele i ich parametry.

Step 1	Step 2	Step 3	Step 4
			
Central server chooses a statistical model to be trained	Central server transmits the initial model to several nodes	Nodes train the model locally with their own data	Central server pools model results and generate one global mode without accessing any data

PODSUMOWANIE

- zainteresowanie łączeniem sieci neuronowych z kryptologią od dawna - nie jest to popularna dziedzina badań, ale ciągle rozwijana
- na razie brak praktycznych zastosowań – tradycyjne podejścia skuteczniejsze i chętniej używane
- nauka modeli na zaszyfrowanych danych może stać się powszechna

- Baluja, Shumeet. "Hiding images in plain sight: Deep steganography." *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017.
- Abadi, Martín, and David G. Andersen. "Learning to protect communications with adversarial neural cryptography." *arXiv preprint arXiv:1610.06918*. 2016.
- Sagar, Vikas, and Krishan Kumar. "Autoencoder Artificial Neural Network Public Key Cryptography in Unsecure Public channel Communication." 2019
- Greydanus, Sam. "Learning the enigma with recurrent neural networks." *arXiv preprint arXiv:1708.07576* 2017.
- Gomez, Aidan N., et al. "Unsupervised cipher cracking using discrete gans." *arXiv preprint arXiv:1801.04883* 2018.
- Gilad-Bachrach, Ran, et al. "Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy." *International Conference on Machine Learning*. PMLR, 2016.
- Gentry, Craig. *A fully homomorphic encryption scheme*. Vol. 20. No. 9. Stanford: Stanford university, 2009.