Disguising Text as an Image using Generative Adversarial Network

Anditya Arifianto
anditya@telkomuniversity.ac.id

Artificial Intelligence Laboratory
School of Computing, Telkom University

Securing Messages

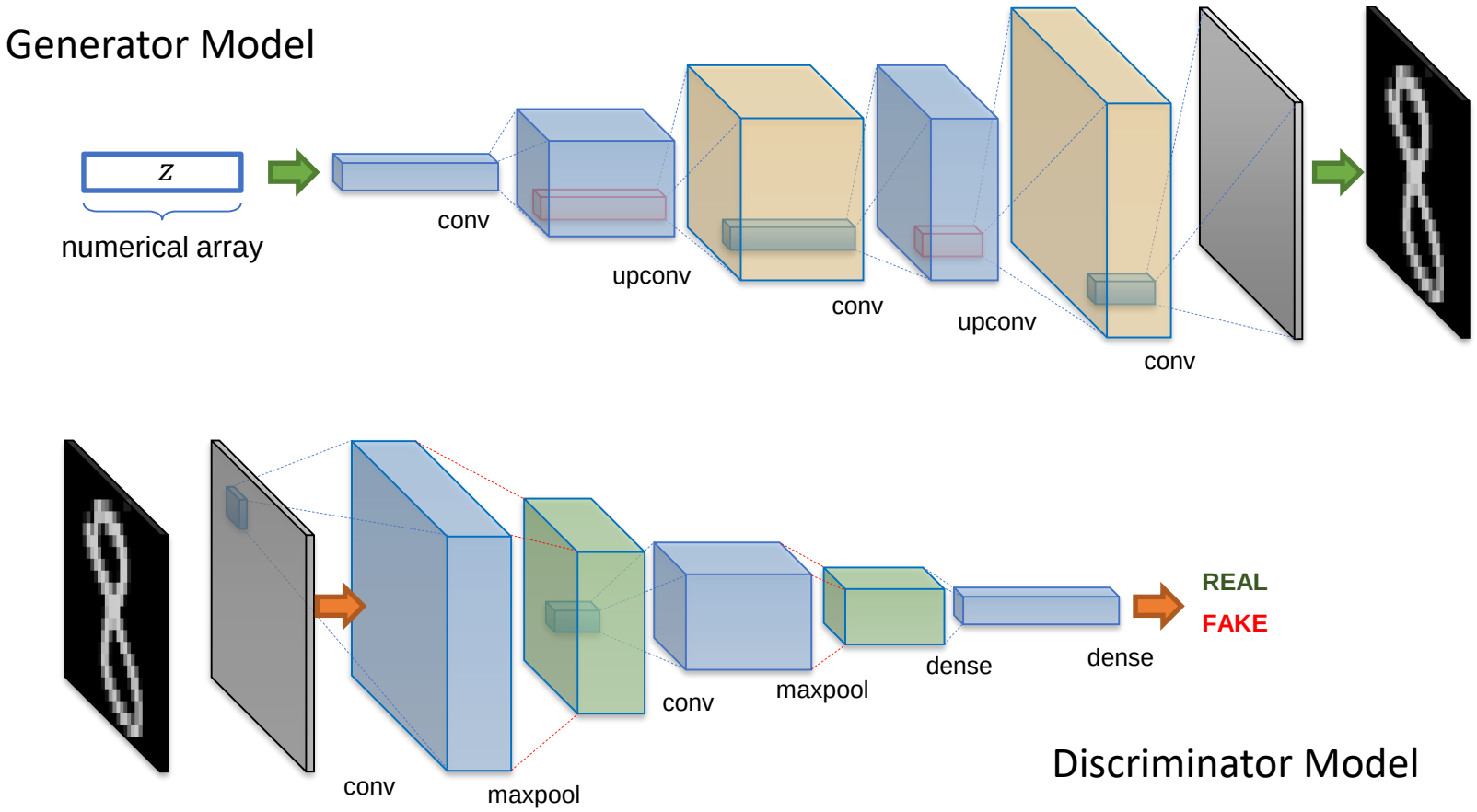
- Era of Internet of Things
  - Everyone and everything is connected
- Digital Information Security efforts
  - Cryptography
  - Steganography
  - Block Chain
  - Etc.

Securing Messages

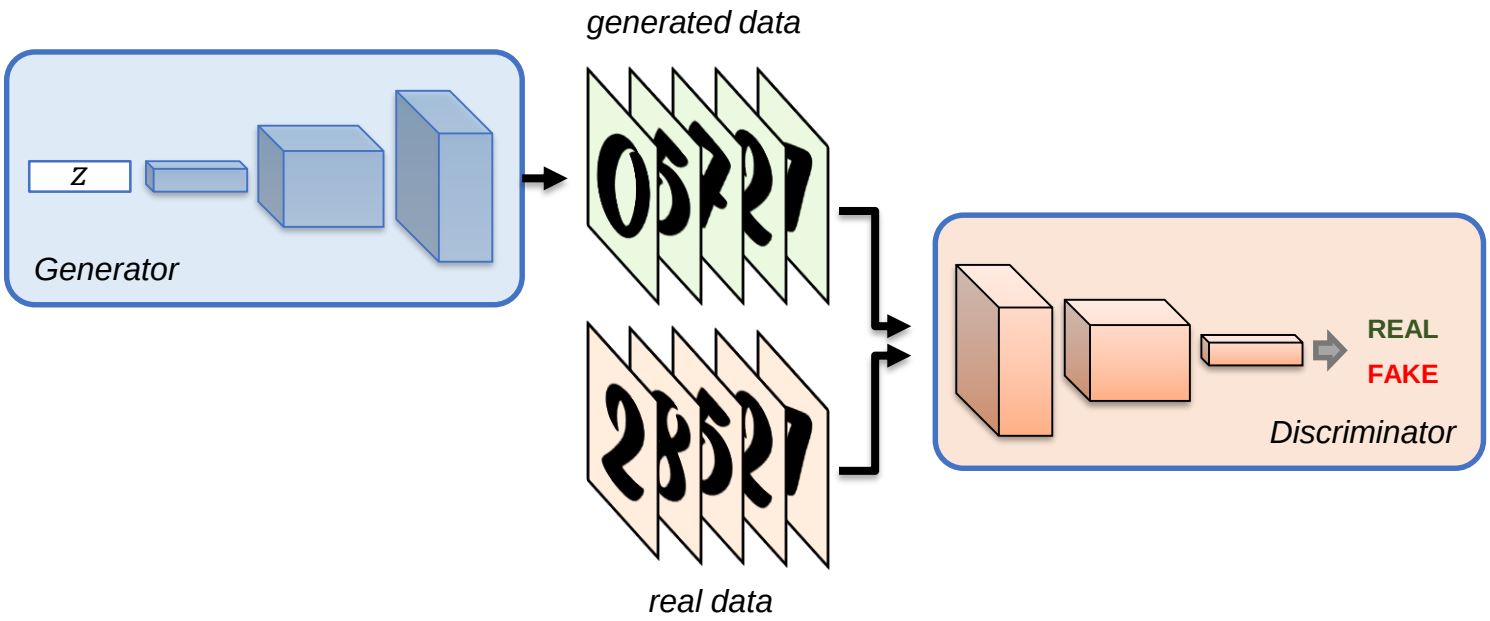
- Indonesian Government is always trying to improve regulation and policy of digital security
  - Electronic Information and Transactions Law
  - Information Security Management System Regulation
  - The formation of National Cyber and Encryption Agency (BSSN)
  - Personal Data Protection bill

- The establishment of regulations must be balanced with improved security methods

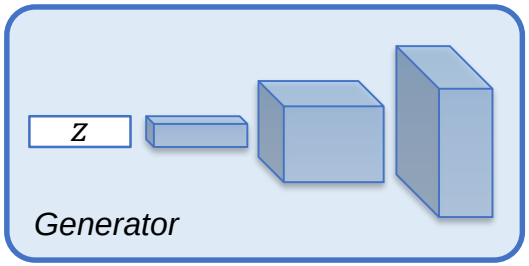
Generative Adversarial Network

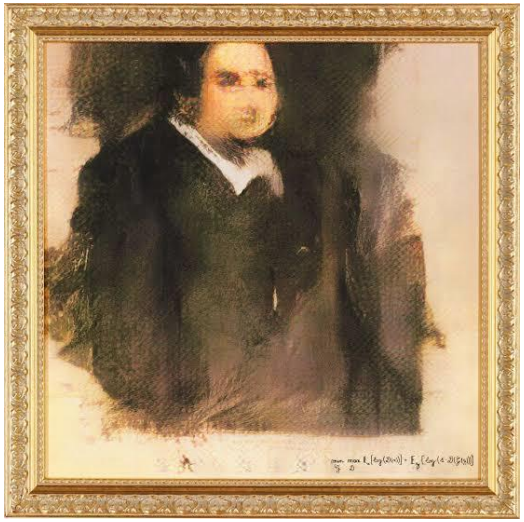


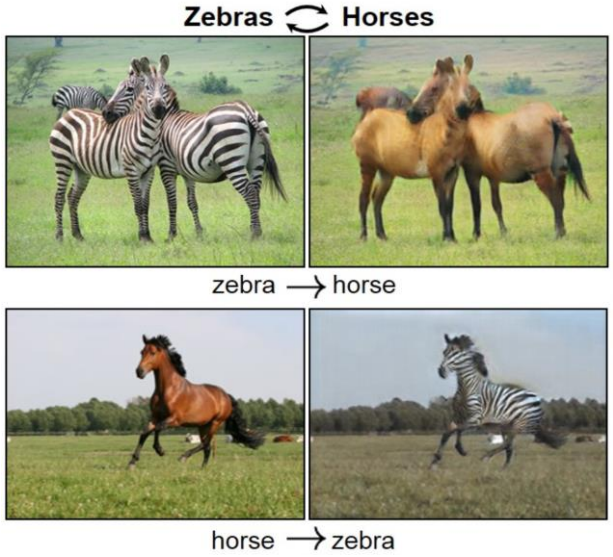
Training GAN

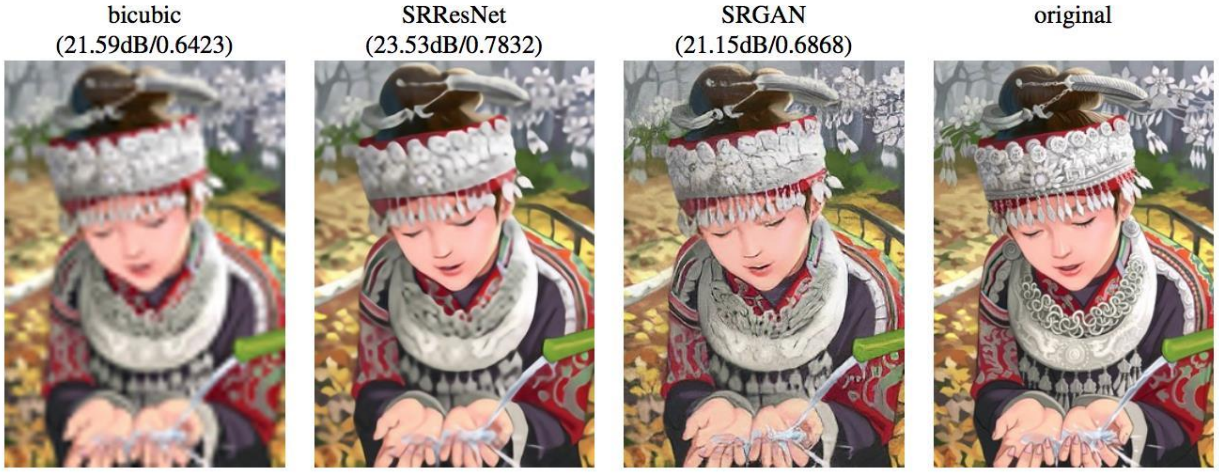


Use the Generator



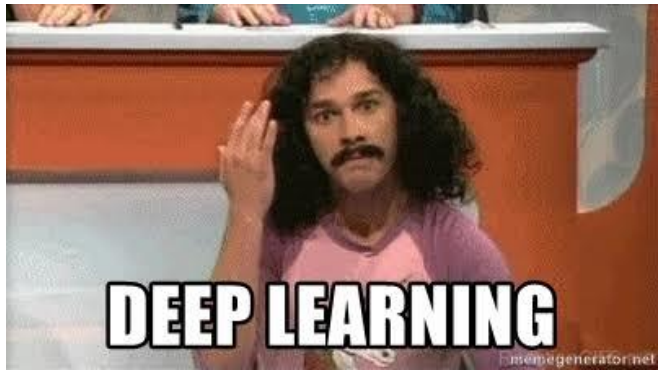


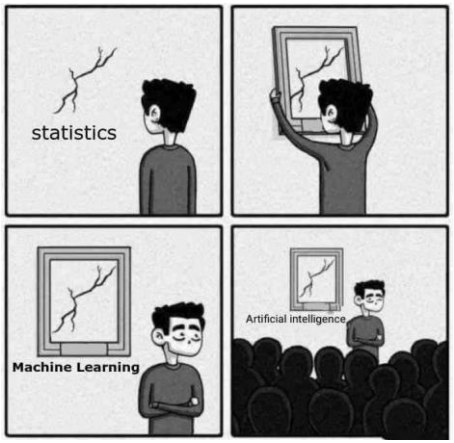




Communicating via Images

- Nowadays, it is very common to communicate by sending images to each other





Messages inside Images

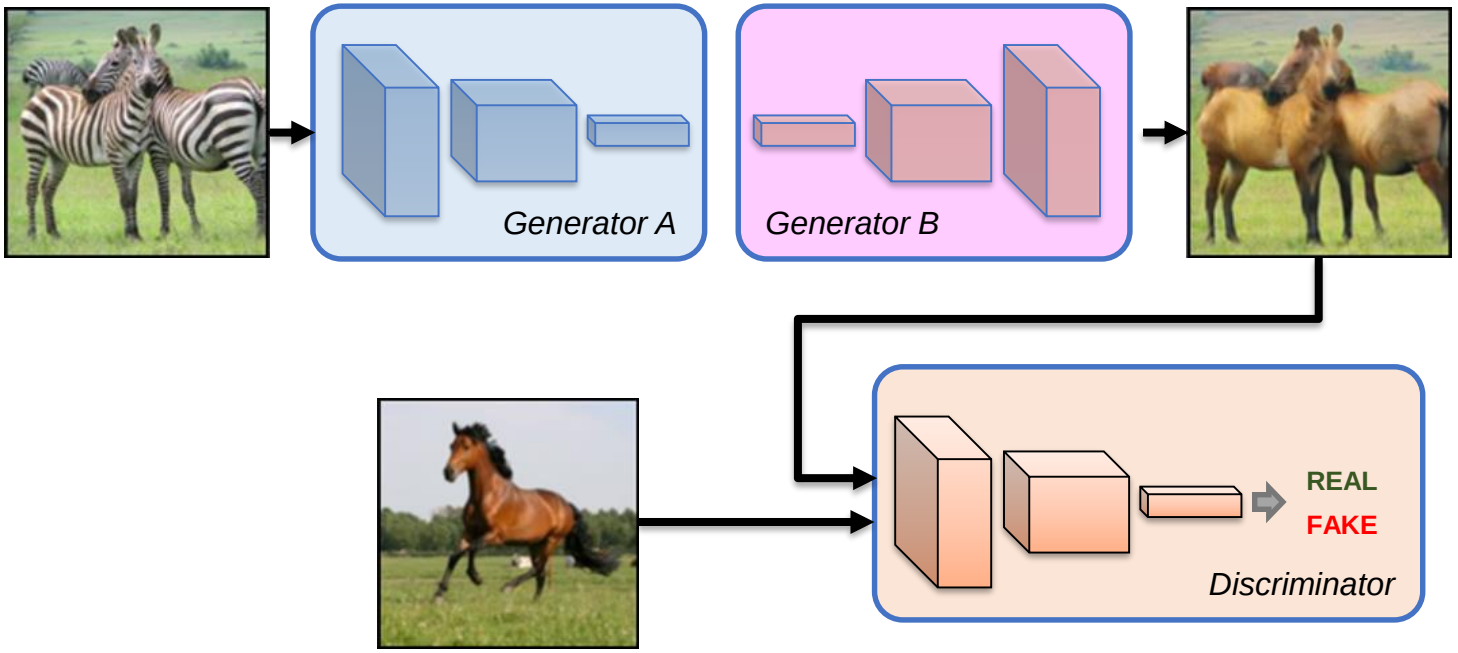
- It's common to hide message inside an image using Steganography techniques.
- Traditional techniques are usually deterministic

- Counter technique: Steganalysis
- Traditional Steganography techniques on Image cover are vulnerable to being detected by Steganalysis

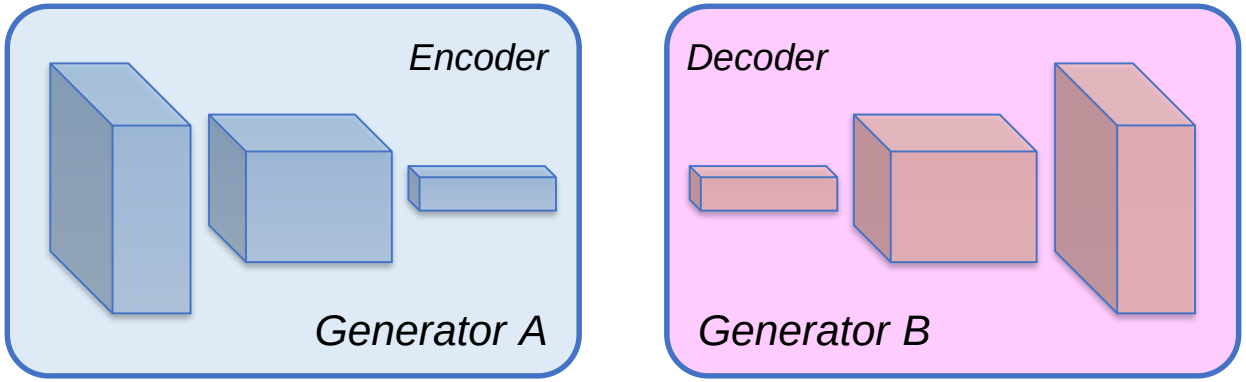
Messages inside Images

- How to create images that have subtle or hidden messages in them that cannot be detected by steganalysis technique?
- How to utilize Deep Learning techniques and use them for Steganography and Cyber Security in general?
- An area that still not widely addressed

CycleGAN

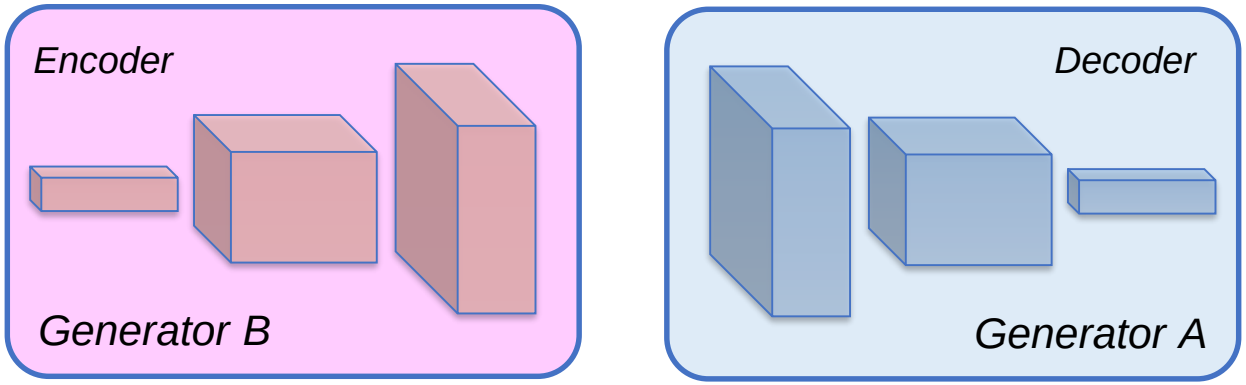


CycleGAN Generator



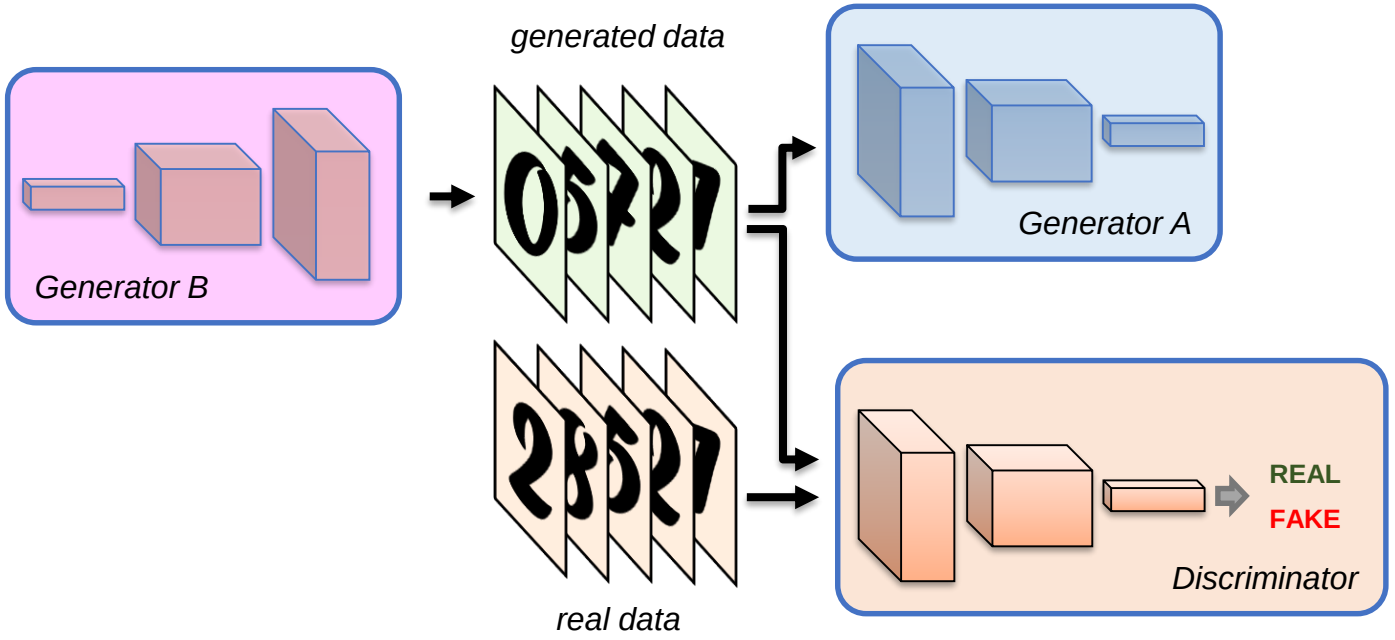
- Image-to-image translation
- A form of Auto Encoder

Reverse Generator

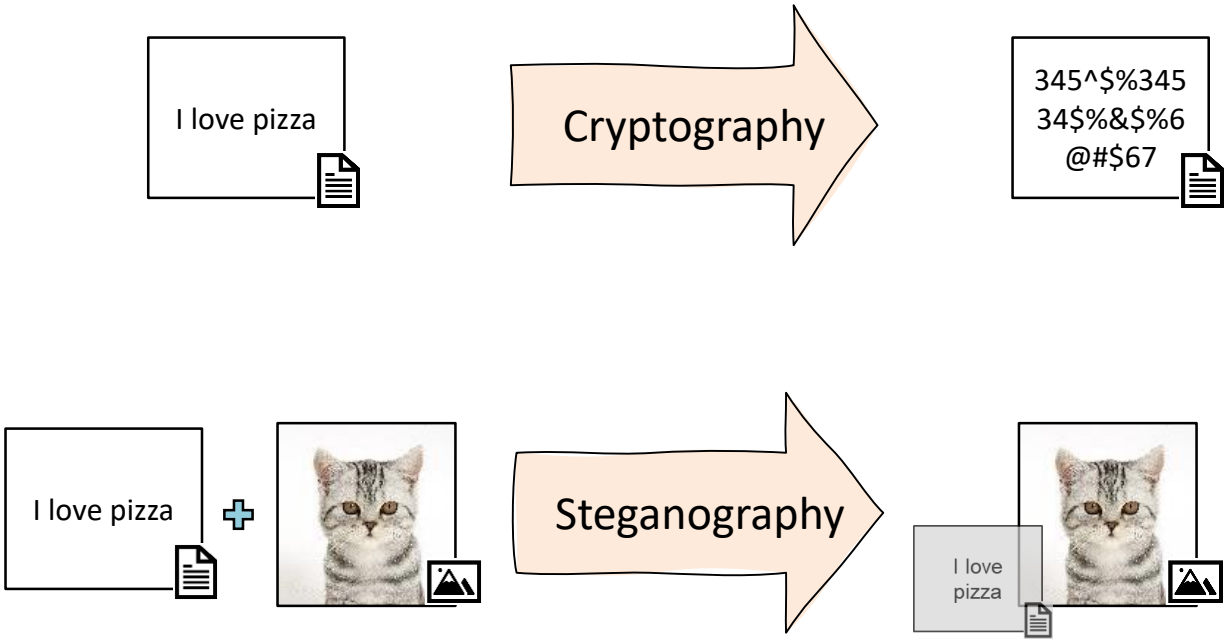


- Switch the generator order
- Vector-to-vector translation

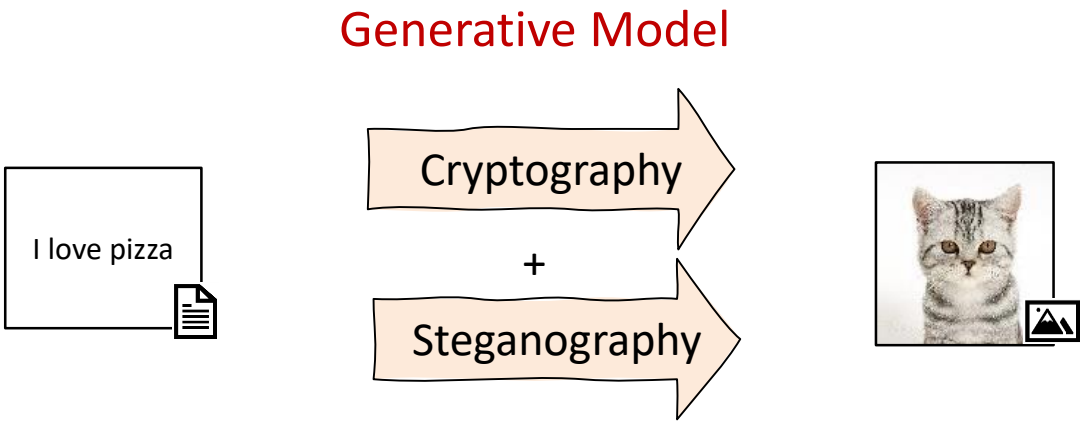
Reverse GAN


generated data
Generator B
Generator A
REAL
FAKE
Discriminator
real data

Cryptography and Steganography

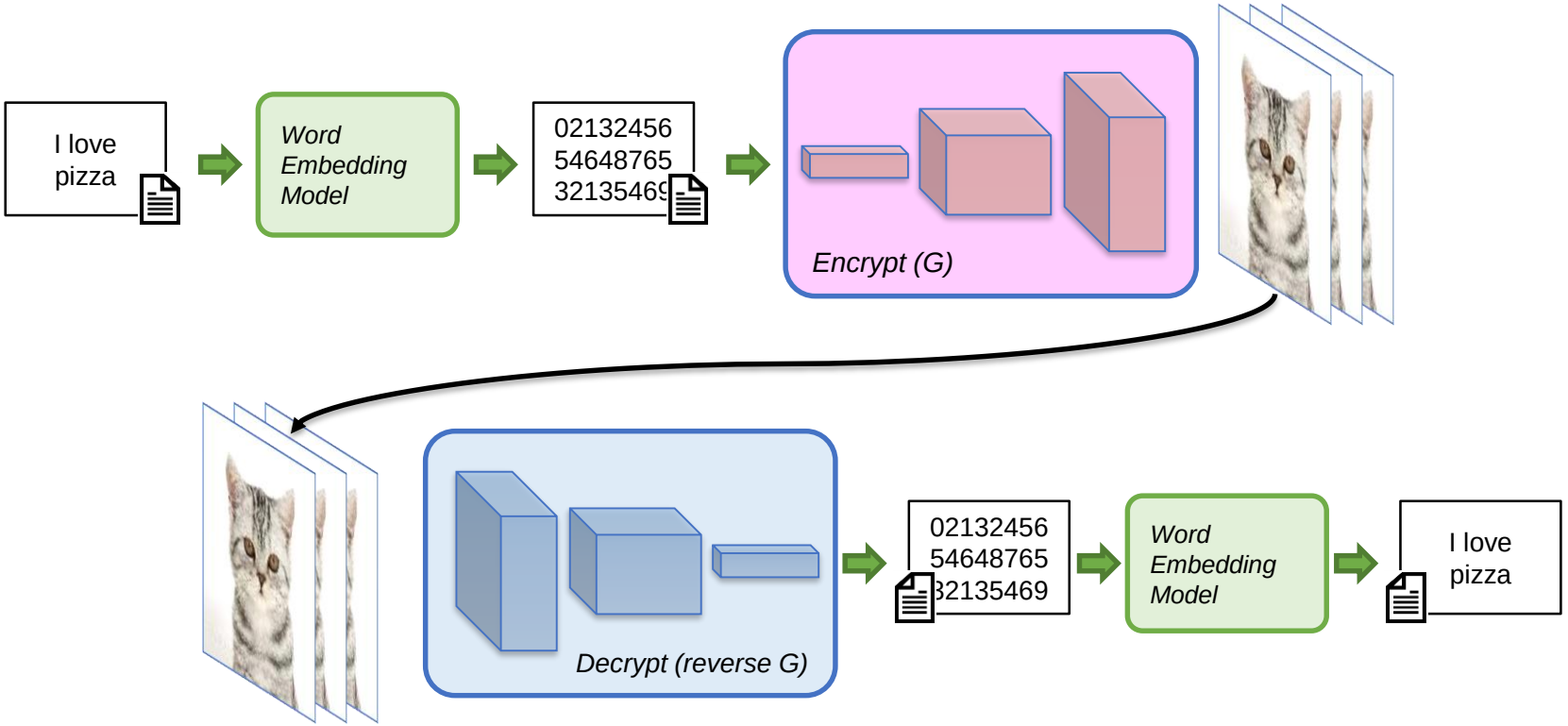


Cryptography and Steganography



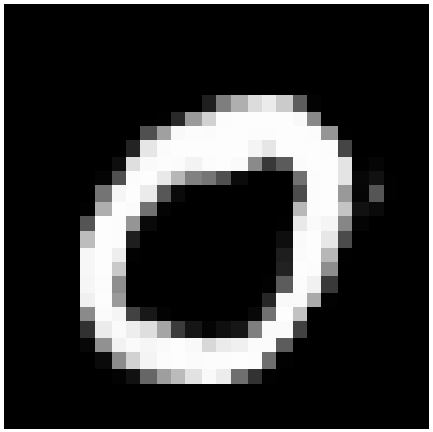
- More robust to steganalysis,
- Images are generated from scratch, not modified
- Images looks natural

Cryptography and Steganography



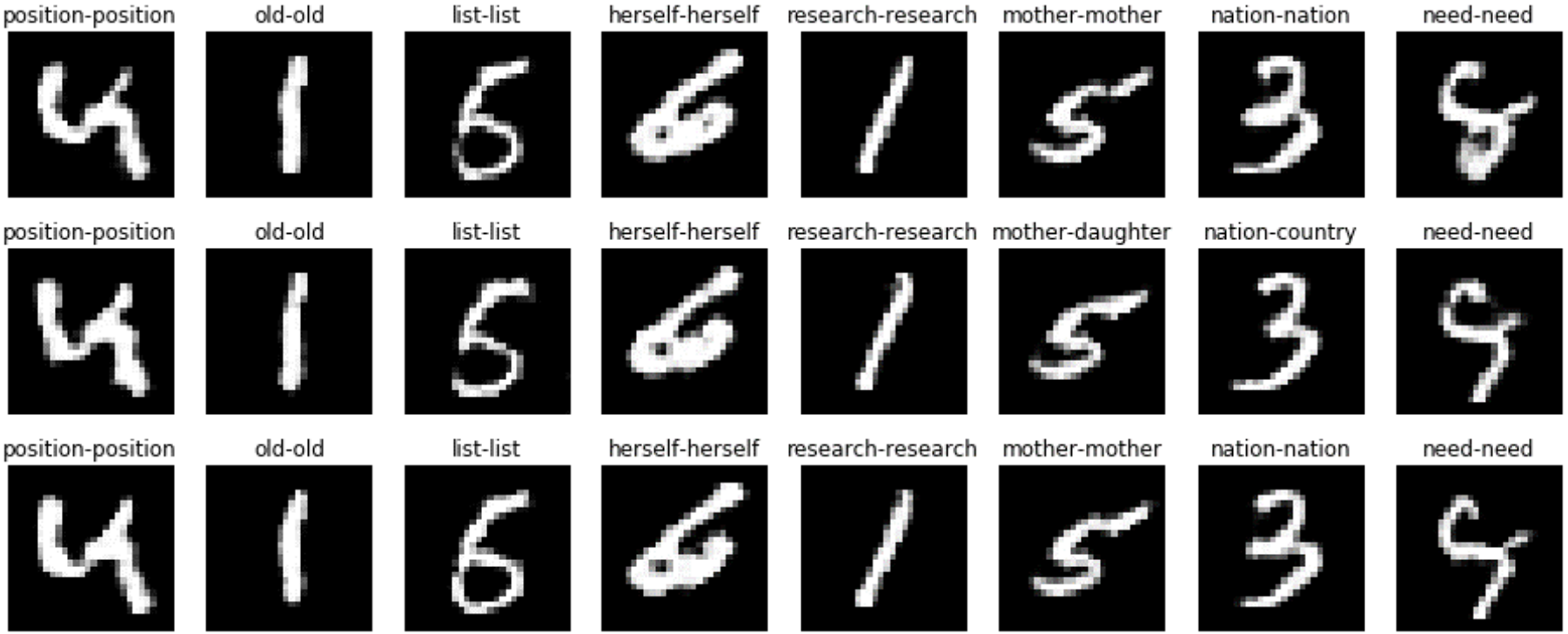
Experiments

- MNIST Image
- Word2Vec Embedding
- 1000 most common English words



Results

- 98.9% accuracy with (-0.1,0.1) noise sample

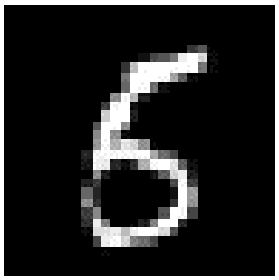


Results

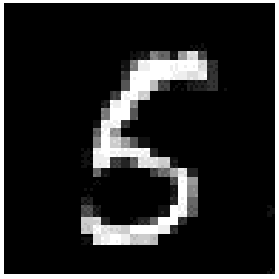
- Generating images from the same word produces slightly different images
- But they are still recognized

- 100% acc w/ Exact generator
- 98.9% acc w/ 0.1 noise
- 80% acc w/ 0.2 noise

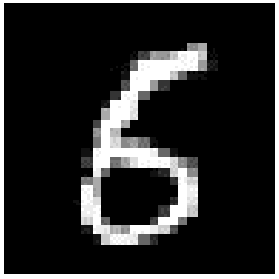
list-list



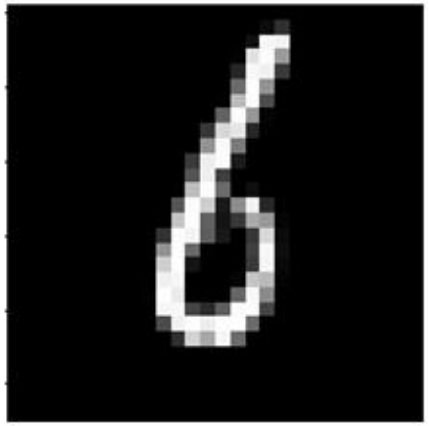
list-list



list-list



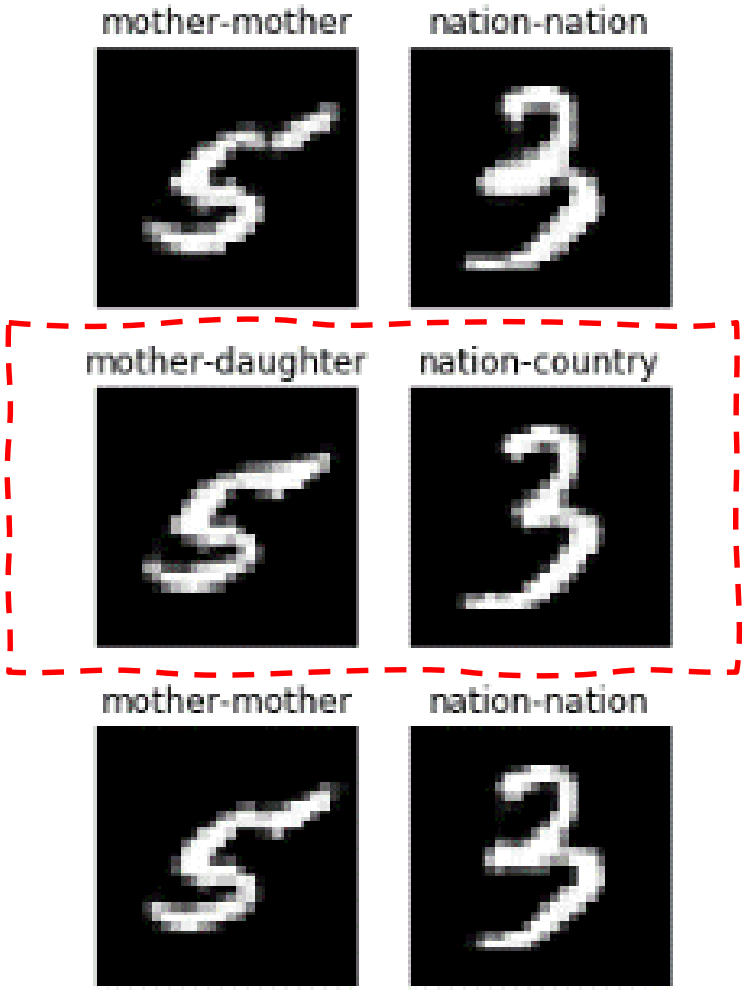
word test: star



```
=== Predicted ===
[0.036488377, 'star']
[2.711475, 'the']
[2.751789, 'who']
[2.7655568, 'but']
[2.7877424, 'he']
[2.794735, 'while']
[2.7952807, 'player']
[2.7991688, 'one']
[2.803216, 'in']
[2.832423, 'just']
```

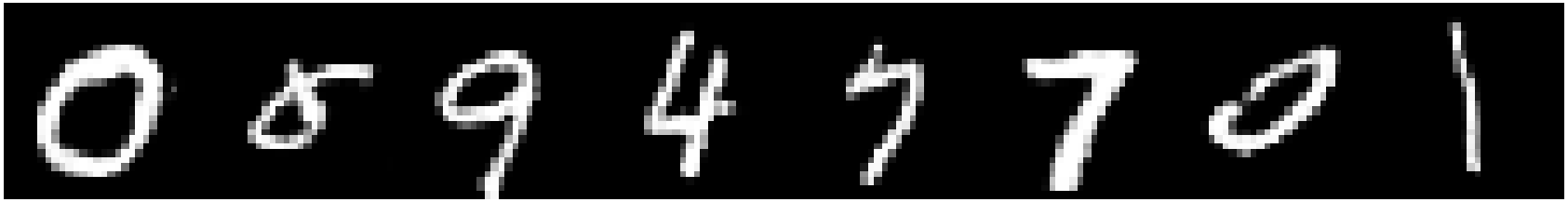
Results

- Although there are still some images that are misidentified



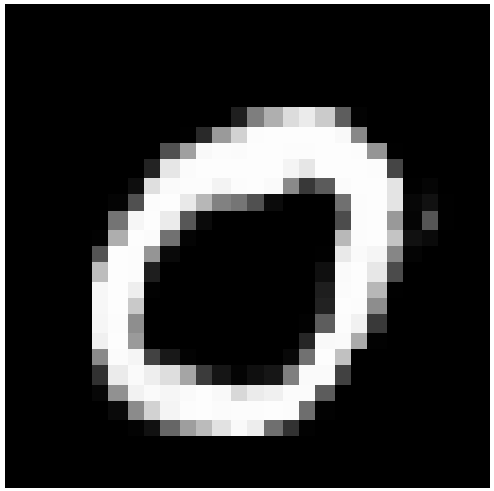
Results

original message: happy boy play music alone outside his house



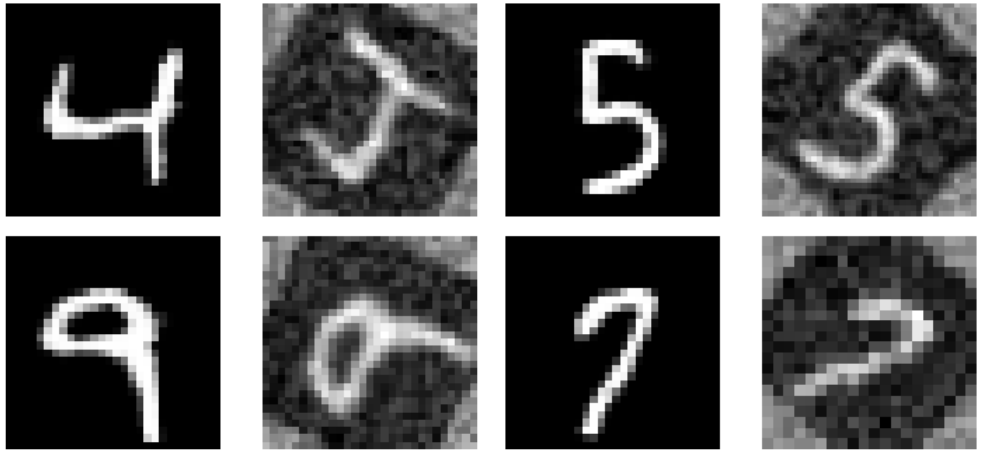
received message: happy boy play music alone outside his house

- Generate interpolation animation
- The message still recognized



Attack Experiments

- Resize Attack
  (0.5,2.0) - 98%
- Rotate Attack
  (-45°, 45°) - 95%
- Noise Attack
  (-0.2,0.2) - 89%



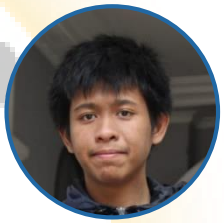
Further Improvements

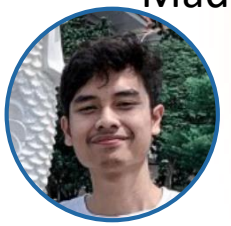
- Private/public key
- Larger image (also color)
- Larger dictionary
- Better embedding model
- Architecture observation
- Input image whole
- Input image parts
- Style Transfer Mechanism

Thank You


Made Raharja SM


Anditya Arifianto


Adriansyah Dwi R


Malik Anhar M


Triwidyastuti J


Rajabandanu S


Muhammad Ferianda S

Contact: anditya@telkomuniversity.ac.id,
https://www.linkedin.com/in/adityaarifianto/