# UNSUPERVISED CIPHER CRACKING USING DISCRETE GANS

Aidan N. Gomez, Sicong Huang, Ivan Zhang, Bryan M. Li, Muhammad Osama, Lukasz Kaiser

January 2018

*Presented by Morten Munk*

# Contents

1. Introduction ..... 2
2. Key Concepts ..... 4
   2.1 Ciphers ..... 5
   2.2 GAN ..... 11
3. Related Work ..... 12
   3.1 CycleGAN ..... 13
4. Ideas & Results ..... 14
   4.1 CipherGAN ..... 15
   4.2 Data ..... 19
   4.3 Results ..... 20
5. Contributions ..... 21
6. Criticism ..... 23
7. Relevance to my project ..... 25

# 1. Introduction

## Premise of this paper

*Can a neural network be trained to deduce withheld ciphers from unaligned text, without the supplementation of preexisting human knowledge?*

## Historical context

- Decryption was human-guided

## Unsupervised learning

- Unpaired bank of ciphertext/plaintext
- No knowledge of vocab frequencies or cipher keys

# 2. Key Concepts

Caesar Shift

**PLAIN:** `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`

**3-SHIFTED:** `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`

**Example**: hello world → khoor zruog

Caesar Shift

**PLAIN:** `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`

**3-SHIFTED:** `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`

**Example**: hello world → khoor zruog

**Weakness**

- Trivial to solve with frequency analysis

Vigenère

**PLAIN:** `hello world`

**KEY:** `soda`

**KEYSTREAM:** `sodas odaso`

**ENCRYPTED:** `zsolg krrdr`

| | 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 |

Src: Vigenere Cipher - Udacity, Youtube

Vigenère

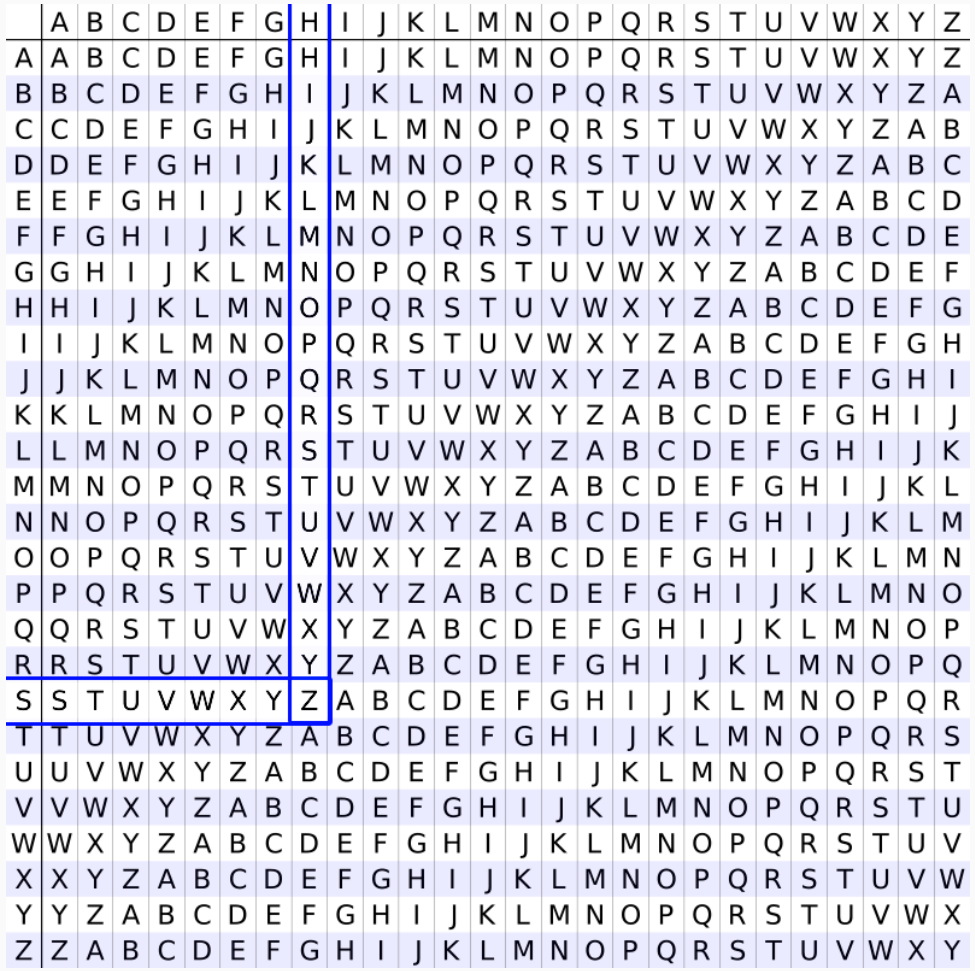
**PLAIN:** `hello world`

**KEY:** `soda`

**KEYSTREAM:** `sodas odaso`

**ENCRYPTED:** `zsolg krrdr`

$(\mathrm{h}, \mathrm{s}) \rightarrow \mathrm{z}$



| | 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 | 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 |

Src: Vigenere Cipher - Udacity, Youtube

Vigenère

**PLAIN:** `hello world`

**KEY:** `soda`

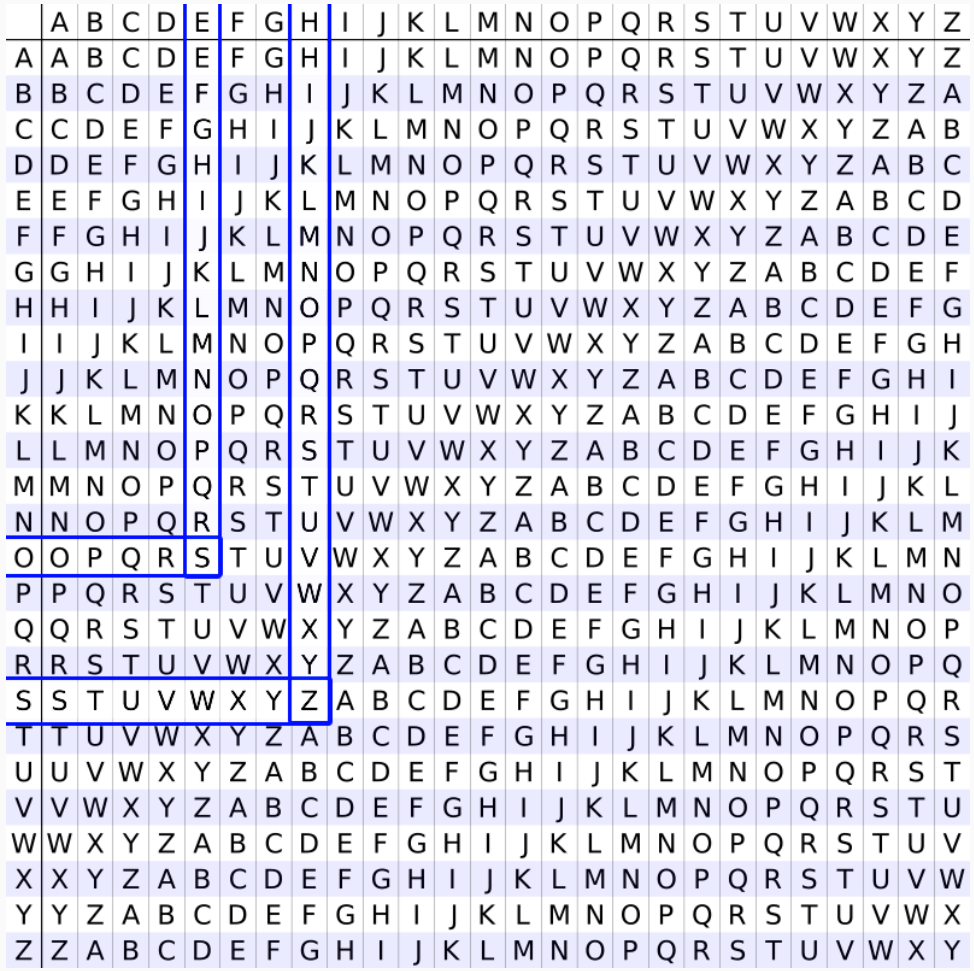
**KEYSTREAM:** `sodas odaso`

**ENCRYPTED:** `zsolg krrdr`

$(\mathrm{h}, \mathrm{s}) \rightarrow \mathrm{z}$

$(\mathrm{e}, \mathrm{o}) \rightarrow \mathrm{s}$

And so on...



| | 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 |

Src: Vigenere Cipher - Udacity, Youtube

Vigenère

**PLAIN:** `hello world`

**KEY:** `soda`

**KEYSTREAM:** `sodas odaso`

**ENCRYPTED:** `zsolg krrdr`

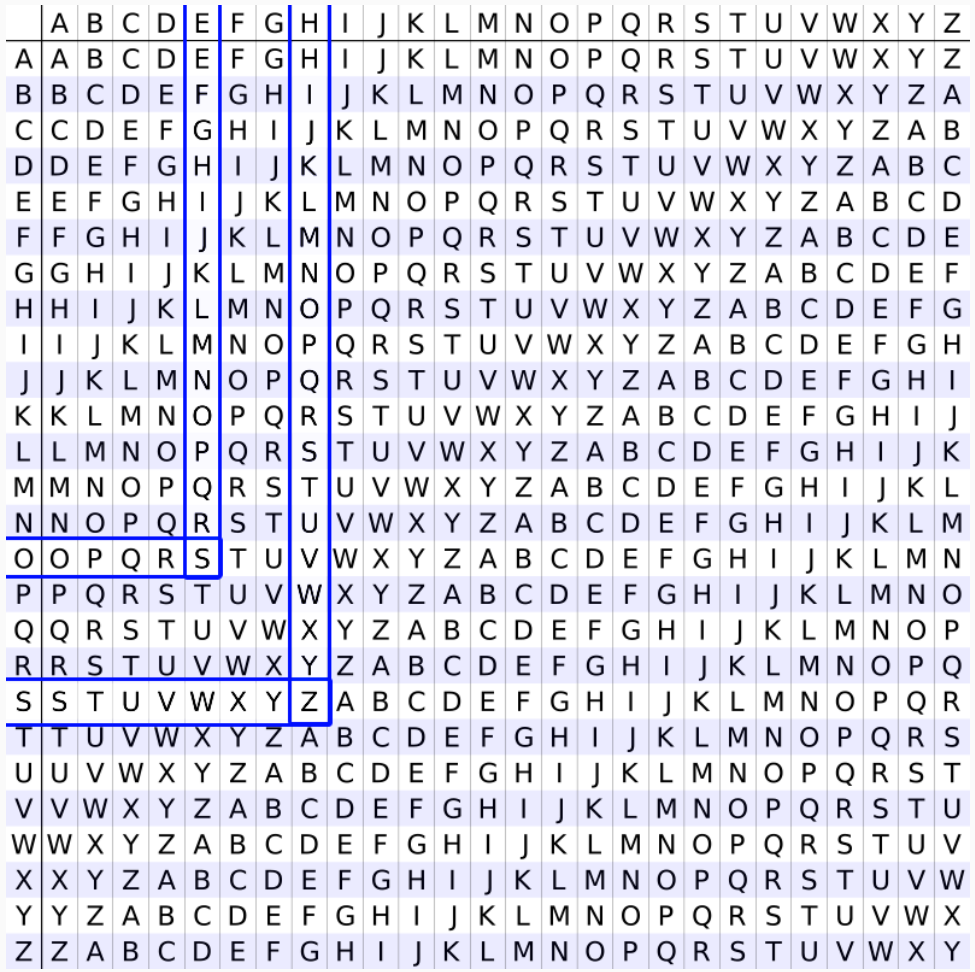
$(\mathrm{h}, \mathrm{s}) \rightarrow \mathrm{z}$

$(\mathrm{e}, \mathrm{o}) \rightarrow \mathrm{s}$

And so on...

**Increased difficulty**

- Frequencies are scrambled



| | 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 |

Src: Vigenere Cipher - Udacity, Youtube

### Generator (G)

- Generates fake samples from noise

### Discriminator (D)

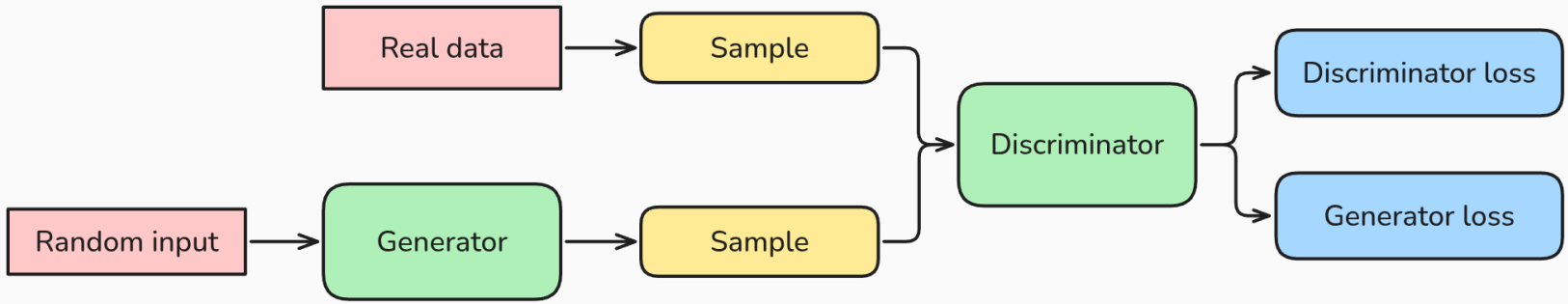
- Tries to classify real/fake

### Minimax game

- D wants to be a good classifier (max)
- G Wants to fool D with fakes (min)

### Convergence

- Similar distribution for real/fake
- D is maximally uncertain (50/50)




Inspired by Goodfellow et al. 2014

## Generator (G)

- Generates fake samples from noise

## Discriminator (D)

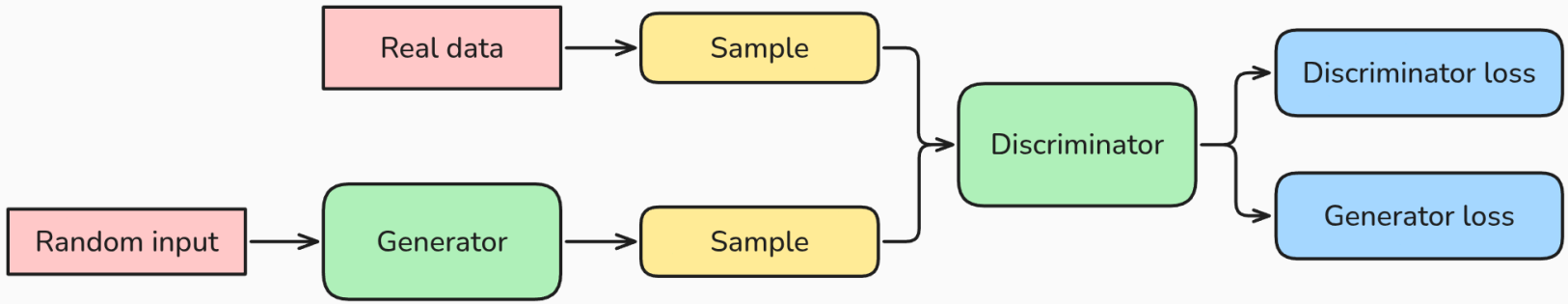
- Tries to classify real/fake

## Minimax game

- D wants to be a good classifier (max)
- G Wants to fool D with fakes (min)

## Convergence

- Similar distribution for real/fake
- D is maximally uncertain (50/50)



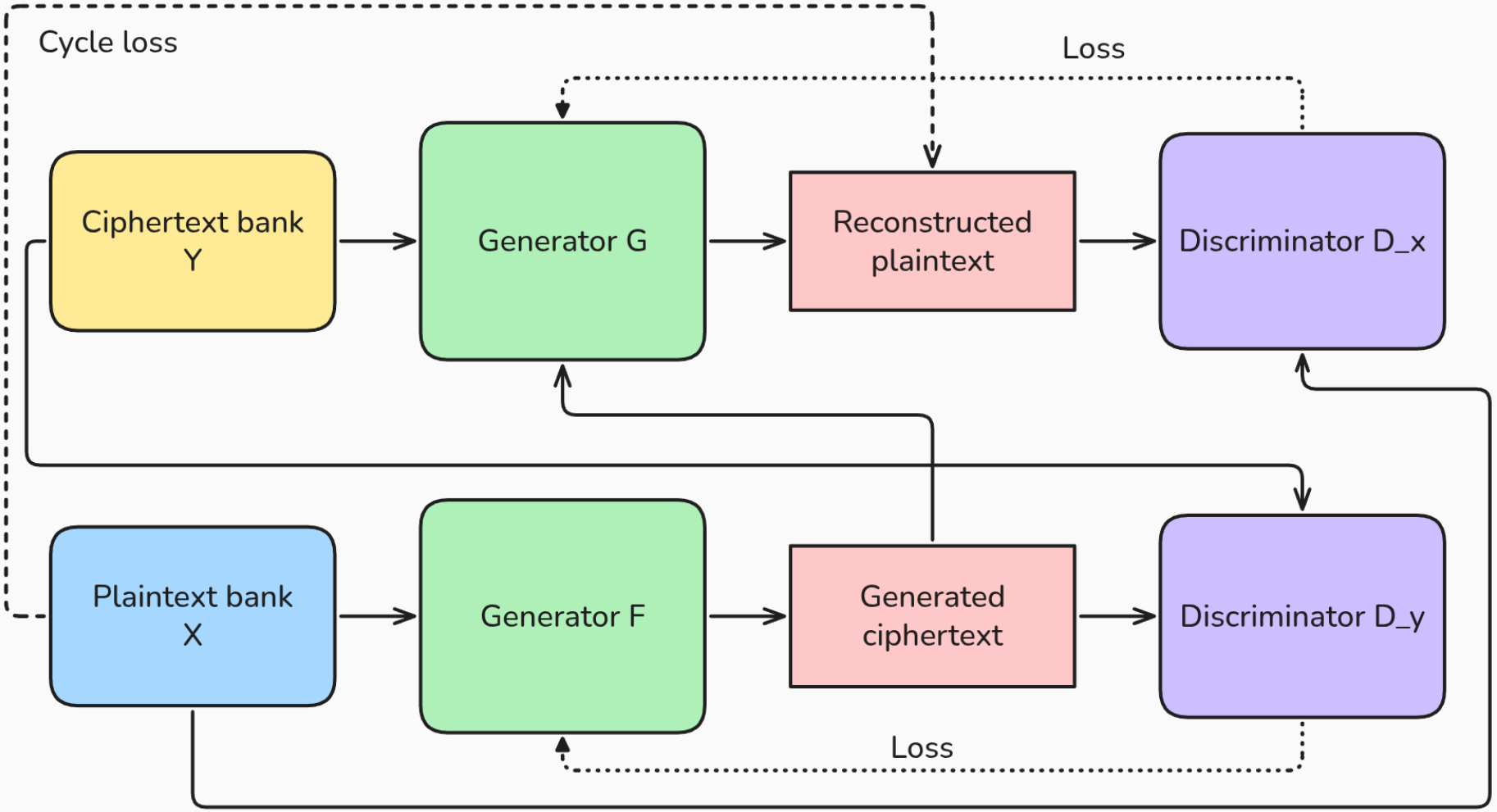

Inspired by Goodfellow et al. 2014

## Vulnerable to mode collapse

- Generator loses diversity - distribution collapses

# 3. Related Work

# 3.1 CycleGAN




Inspired by: Zhu et al., 2017 & Yi et al., 2017 & Liu et al., 2017

# 4. Ideas & Results

Figure 2: Discriminators trained on the toy example of recognizing the bottom-right corner of a simplex as true data. From left to right the discriminators were regularized using: nothing; WGAN Jacobian norm regularization; and, the relaxed sampling technique.

Gomez, et al. 2018

## Original GAN can be too strict

- No helpful feedback

## Used by this paper

- Wasserstein Jacobian norm (Arjovsky et al., 2017 & Gulrajani et al., 2017)
- Relaxation Sampling

## Words in embedding space

- Discrete choices do not produce gradients

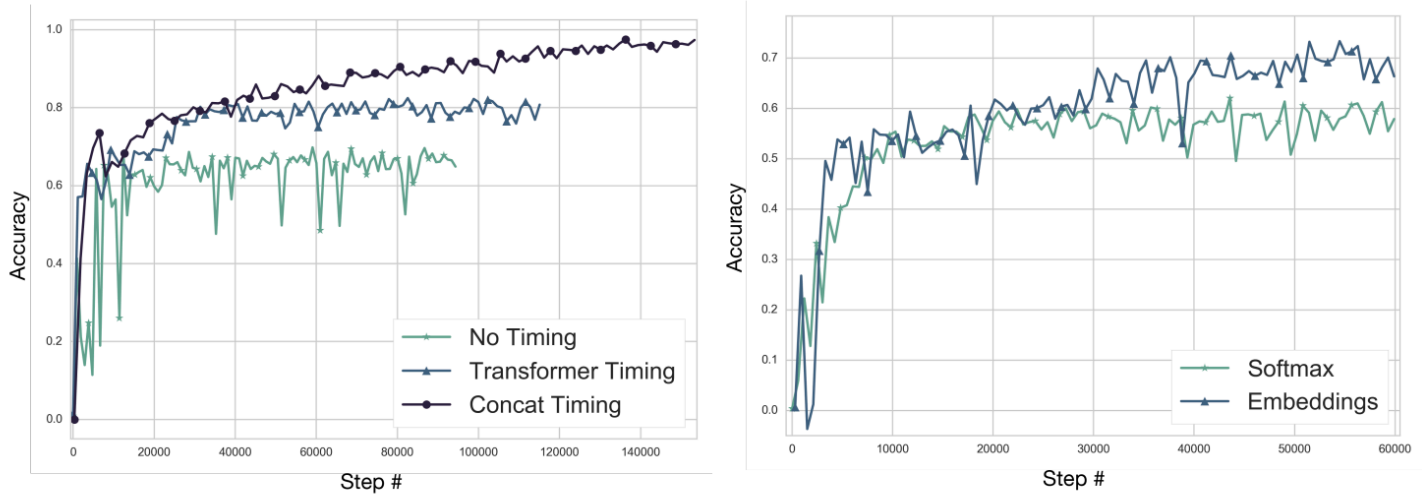



Figure 3: Left: Comparison of different timing techniques for Brown-C Vigenère. Right: Comparison of embedding vs. raw softmax on Brown-W with vocab size of 200.

Gomez, et al. 2018

## Positional embedding (Timing)

- Vigenère relies on positioning
- Position is described using vector in embedding

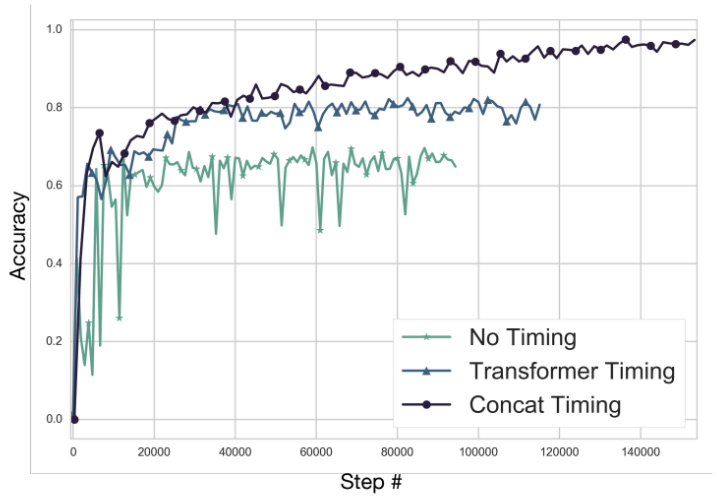
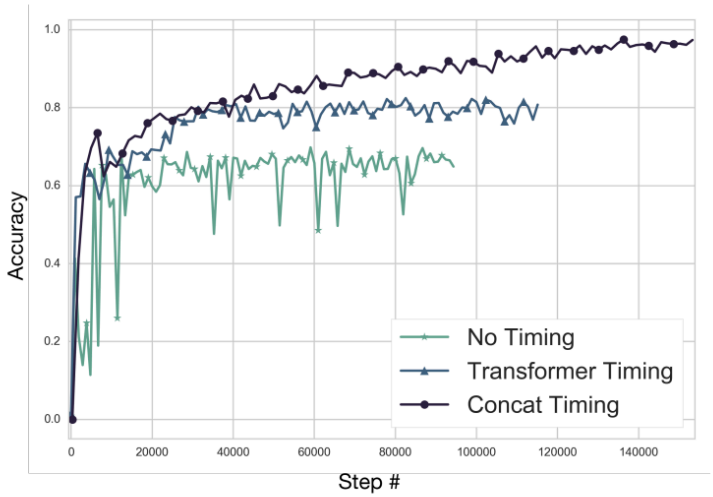



Figure 3: Left: Comparison of different timing techniques for Brown-C Vigenère. Right: Comparison of embedding vs. raw softmax on Brown-W with vocab size of 200.

Gomez, et al. 2018

## Squared difference instead of log likelihood

- Provided stable gradient
- Determines distance from target



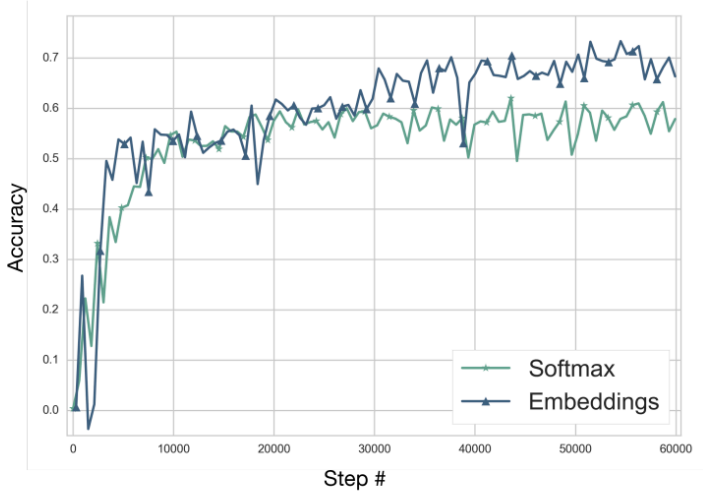




Figure 3: Left: Comparison of different timing techniques for Brown-C Vigenère. Right: Comparison of embedding vs. raw softmax on Brown-W with vocab size of 200.

Gomez, et al. 2018

**Brown English text dataset**

- 1st half as plaintext $\mathcal{X}$ distribution
- 2nd half as ciphertext $\mathcal{Y}$ distribution

**Brown-W**

- Top 200 most frequent words
- Rest is blank

**Brown-C**

- chars, special letters, punctuation
- 58 elements in vocab

# 4.3 Results

| **Data** | Brown-W | Brown-W | Brown-C | Freq. Analysis (With Key) | |
|---|---|---|---|---|---|
| **Vocab size** | 10 | 200 | 58 | 58 | 200 |
| Cipher | **Shift/Permutation** | | | | |
| Acc. | 100% | 98.7% | 99.8% | 80.9% | 44.5% |
| Cipher | **Vigenère (Key: "345")** | | | | |
| Acc. | 99.7% | 75.7% | 99.0% | 9.6% (78.1%) | <0.1% (44.3%) |

Table 2: Average proportion of characters correctly mapped in a given sequence. The "Freq. Analysis" column is simple frequency analysis applied to the same corpus our model observes. For Vigenère we also show the score if the key were known (note: the key is left unknown to our model).

Gomez, et al. 2018

# 5. Contributions

## CipherGAN Novelty

- First GAN for cryptanalysis
- Continuous relaxation
- WGAN jacobian
- Concat timing

## Performance & scaling

- Accuracy
- Unsupervised
- Larger vocab

| Work | Ciphertext Length | Accuracy |
|---|---|---|
| Hasinoff (2003) | 500 | $\sim 97\%$ |
| Forsyth & Safavi-Naini (1993) | 5000 | $\sim 100\%$ |
| Ramesh et al. (1993) | 160 | $\sim 78.5\%$ |
| Verma et al. (2007) | 1000 | $\sim 87\%$ |

Table 1: Previous results on automated shift cipher cracking with limited ciphertext length.

Gomez, et al. 2018

# 6. Criticism

# 6. Criticism

**Why not test on a real cipher?**

- E.g. Civil War Vigenere cipher
- Real ciphers has typos, etc.

**Cipher lengths**

- No real stress testing

**Missing other neural baselines**

- LSTM or Transformers

**Missing sections**

- Related work
- Future Work

# 7. Relevance to my project

# 7. Relevance to my project

## Homophonic Substitution Ciphers

- 1-to-many mappings
- English
- Only lowercase letters
- No spaces

## Comparative Study

- Embeddings
- Heuristics
- Seq-2-Seq

## Increase sequence length with CausalLM

- Linear attention?
- Flash attention?

**Decipherment as Regression: Solving Historical Substitution Ciphers by Learning Symbol Recurrence Relations**

**Nishant Kambhatla** **Logan Born** **Anoop Sarkar**
School of Computing Science, Simon Fraser University
8888 University Drive, Burnaby BC, Canada
{nkambhat, loborn, anoop}@sfu.ca

**Abstract**

Solving substitution ciphers involves mapping sequences of cipher symbols to fluent text in a target language. This has conventionally been formulated as a search problem, to find the decipherment key using a character-level language model to constrain the search space. This work instead frames decipherment as a sequence prediction task, using a Transformer-based causal language model to learn recurrences between

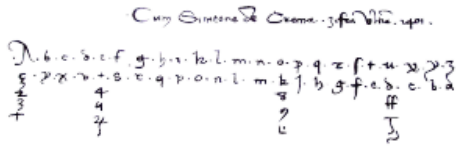


Figure 1: The homophonic substitution key for the *Simeone de Crema* written in Mantua in 1401 AD. The top line maps each character in the alphabet to its reversed-alphabet equivalent; each vowel is substituted by three additional symbols.

# 7. Relevance to my project

**Transpositions**

- Maybe concat timing?

**CipherGAN will likely struggle**

- HSC is not uniquely invertible
- Maybe GAN is too unstable for HSC

**Supervised learning**

- We did not explore unsupervised