Cracking Arbitrary Substitution Ciphers

Xiangxi Mu, Tianze Hua, Yiyang Zhang, Gary Zheng

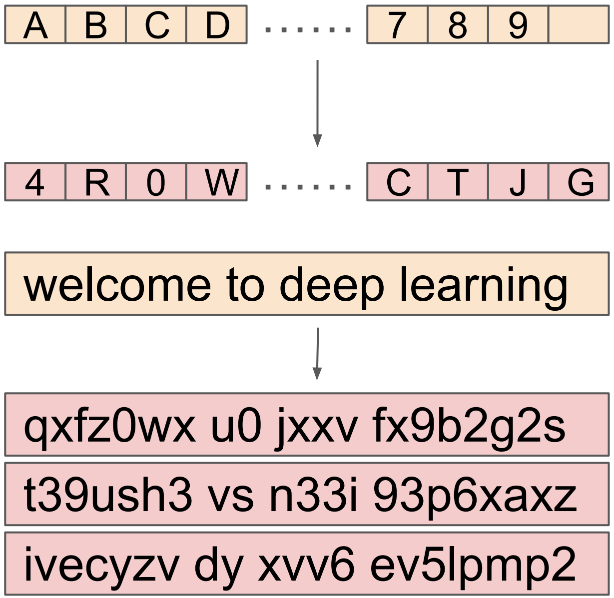
BROWN

Research Question

Can we use transformer models to map ciphertexts back to plaintexts?

Substitution Cipher

```
Subst_cipher

__init__(self, key=None, domain=List):
p2c: dict
c2p: dict
Initializes the cipher maps with randomly
shuffled keys

encrypt(self, plaintext: str) -> List:
Encrypts the plaintext

decrypt(self, ciphertext: str) -> List:
Decrypts the ciphertext
```



Training Data

wikitext-103-raw-v1 for our training, validation, and testing data

- Raw data language corpus extracted from selected Wikipedia articles

Preprocessing:

- letter(switched to lowercase), number, and space
- Concatenate all entries into one string
- Randomly select substrings for training, validation, and testing

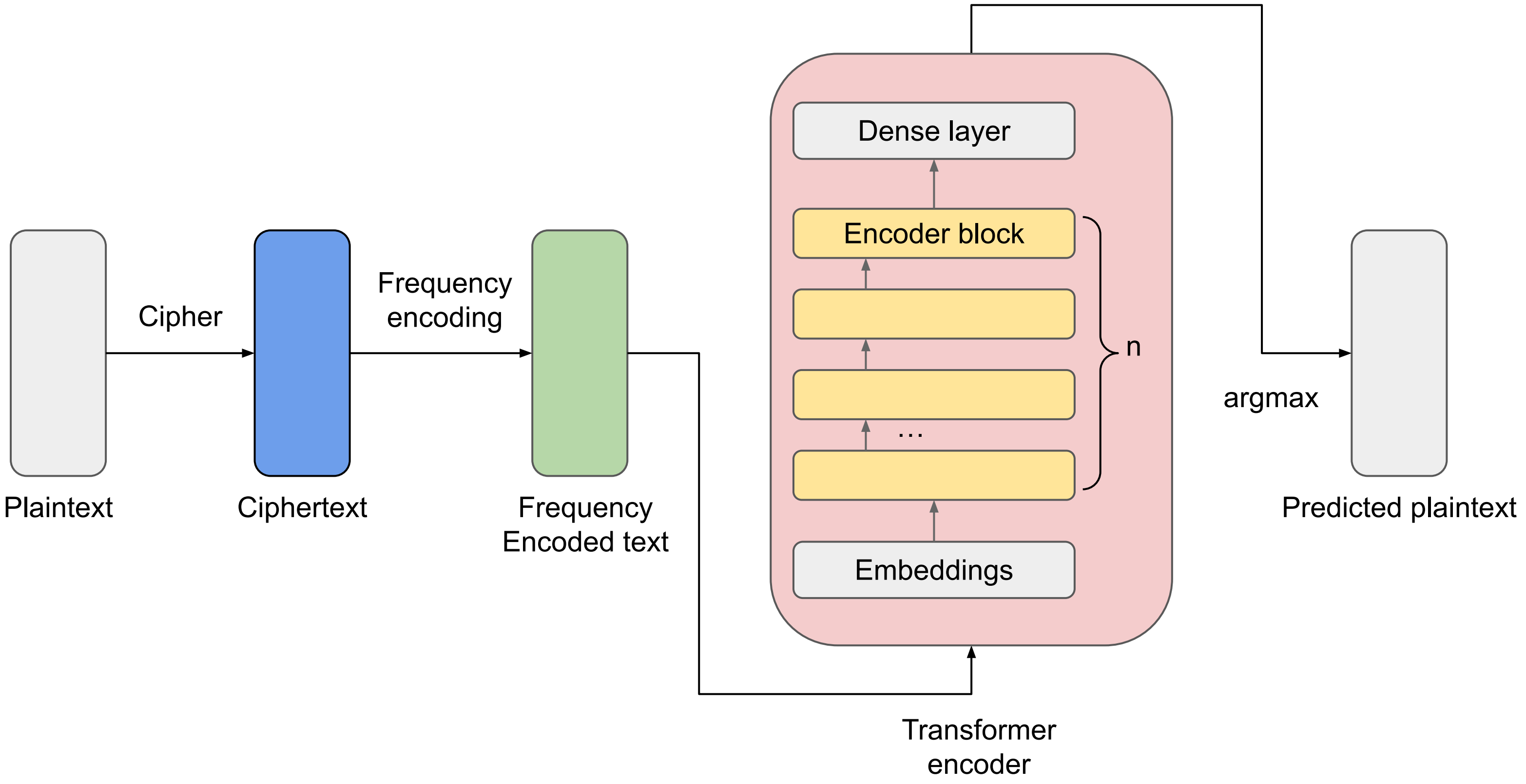
Dataset splits	**Number of characters**
Train	500,978,592
Validation	1,065,888
Test	1,198,238

Frequency Encoding

In a given text, letters and letter combinations (n-grams) appear in varying frequencies, and the character frequency distribution is roughly preserved in any sample drawn from a given language.

To encode that information, we re-map each ciphertext character to a value based on its frequency rank. This way, we convert any cipher text to a frequency-encoded cipher. Intuitively, by frequency encoding, we are reducing the number of possible substitution keys.

Workflow visualization



Plaintext

…beef produced last month by cargill meat solutions were shipped to walmart locations nationwide according to a recall notice…

Ciphertext

…x77b 48k9yg79 snjr 0kmrt xf gn8lqss 07nr jksyrqkmj 5787 jtq4479 rk 5ns0n8r skgnrqkmj mnrqkm5q97 nggk89qml rk n 87gnss mkrqg7…

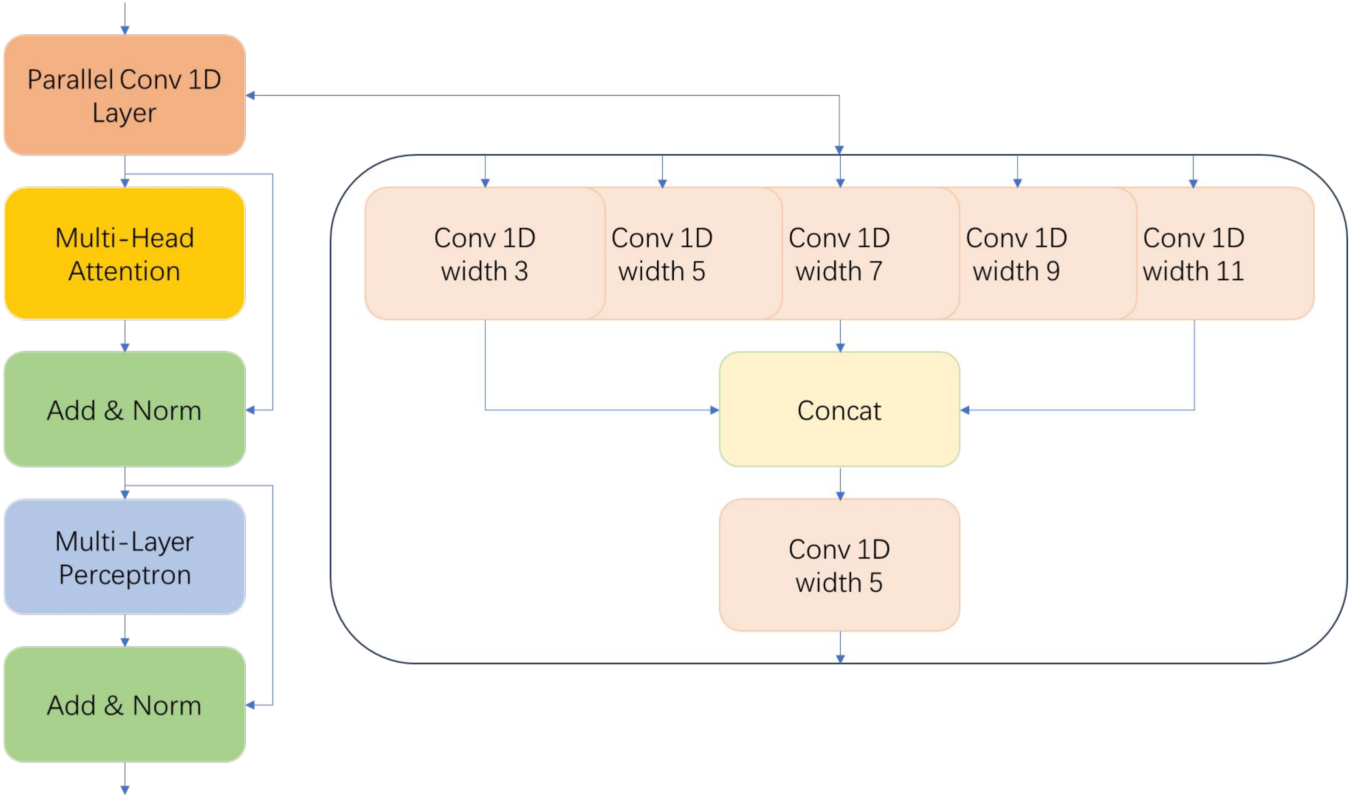
Frequency-Encoded Ciphertext

18, 1, 1, 15, 0, 17, 7, 6, 10, 14, 9, 1, 10, 0, 11, 3, 8, 2, 0, 13, 6, 5, 2, 12, 0, 18, 19, 0, 9, 3, 7, 16, 4, 11, 11, 0, 13, 1, 3, 2, 0, 8, 6, 11, 14, 2, 4, 6, 5, 8, 0, 20, 1, 7, 1, 0, 8, 12, 4, 17, 17, 1, 10, 0, 2, 6, 0, 20, 3, 11, 13, 3, 7, 2, 0, 11, 6, 9, 3, 2, 4, 6, 5, 8, 0, 5, 3, 2, 4, 6, 5, 20, 4, 10, 1, 0, 3, 9, 9, 6, 7, 10, 4, 5, 16, 0, 2, 6, 0, 3, 0, 7, 1, 9, 3, 11, 11, 0, 5, 6, 2, 4, 9, 1

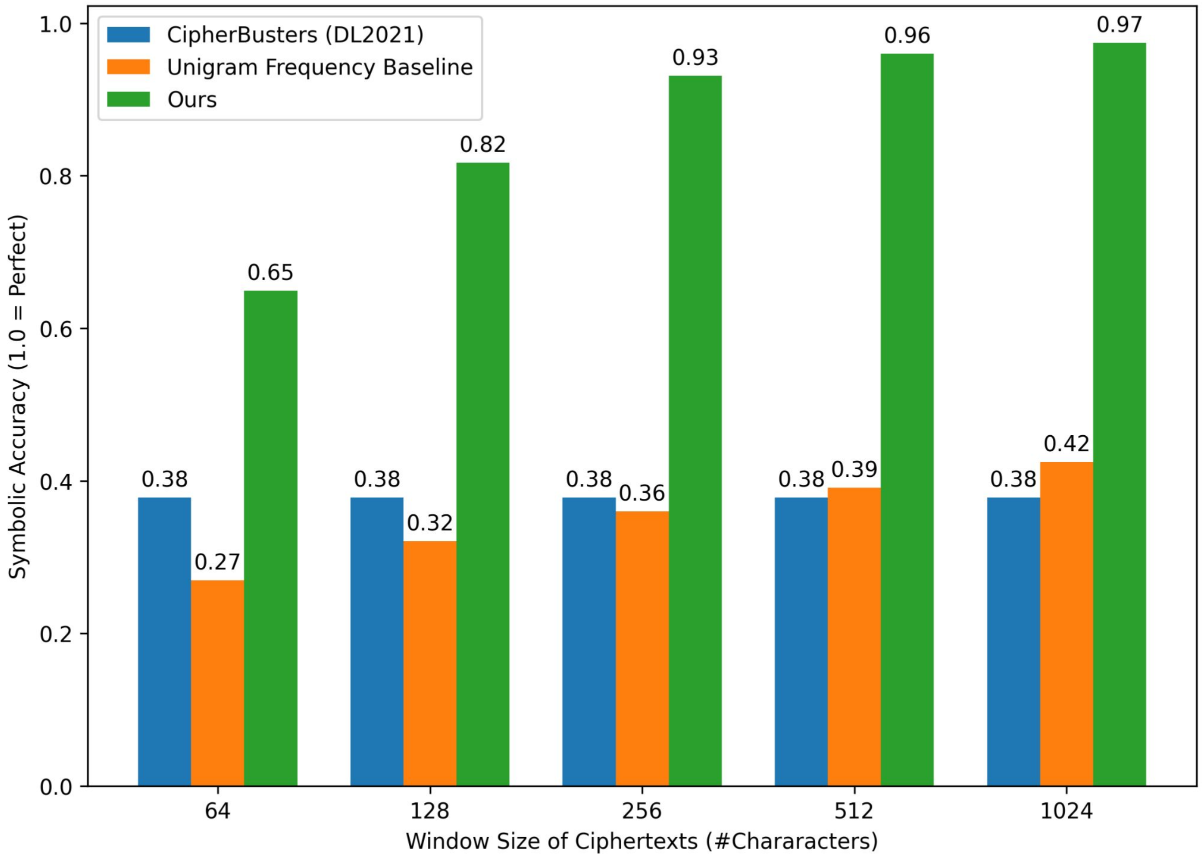
Deciphered Text

…beef produced last month by cargill meat solutions were shipped to kalmart locations nationwide according to a recall notice…

CNN-Enhanced Encoder Blocks



Model Performance Compared to Baselines

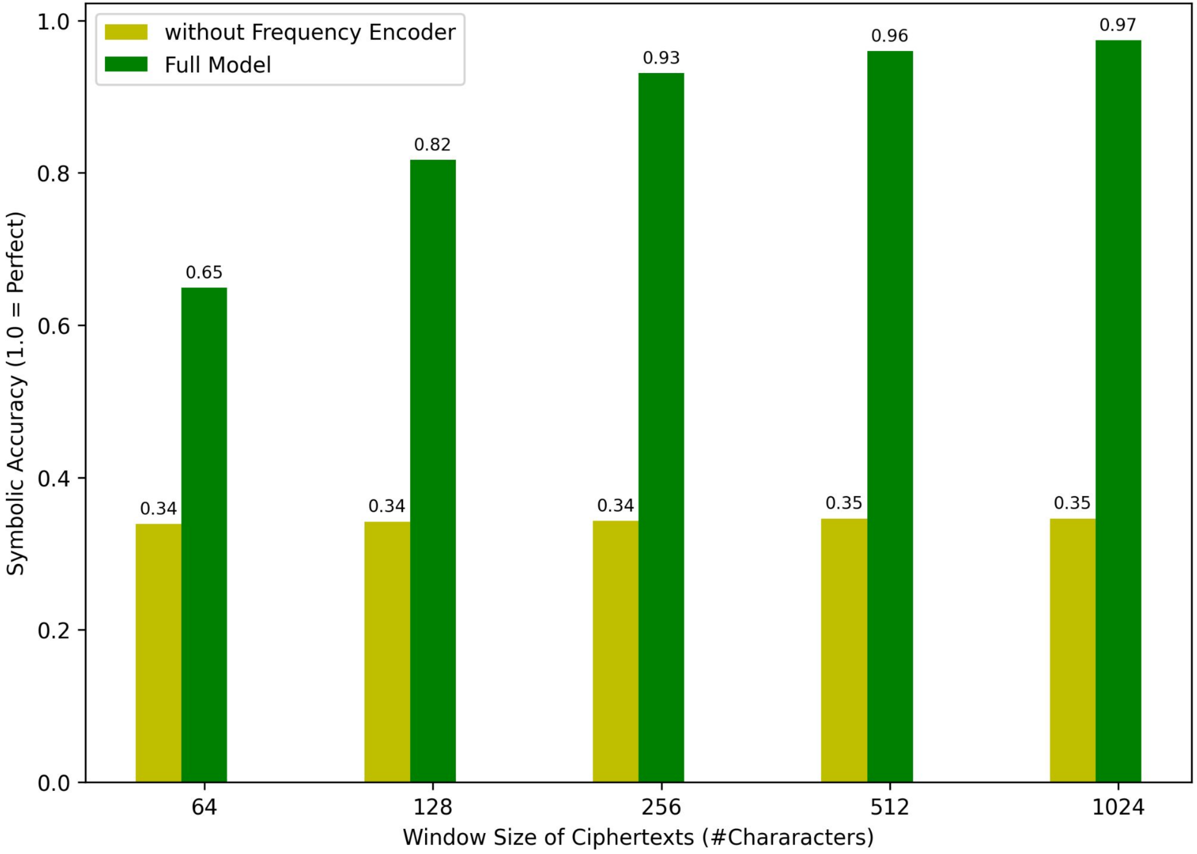


Training/Testing Models on Different Window Sizes

Test/Train	64	128	256	512	1024
64	**0.649**	0.616	0.599	0.521	0.495
128	-	**0.817**	0.814	0.743	0.71
256	-	-	**0.931**	0.89	0.862
512	-	-	-	**0.96**	0.948
1024	-	-	-	-	**0.974**

Table 1: Decipherment performance of models trained with different ciphertext lengths, on ciphertexts of different lengths.

Ablation Studies (Removing Frequency Encoder)

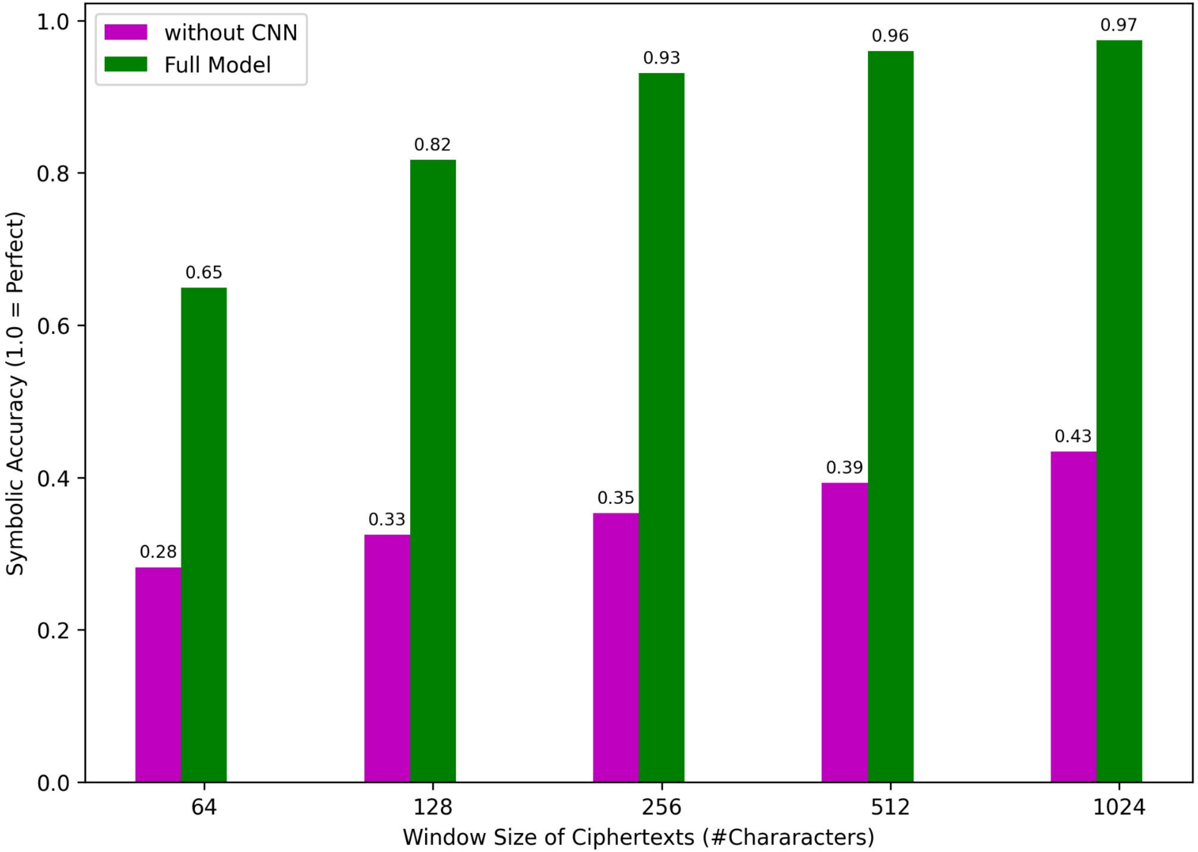


Ablation Studies (Removing Frequency Encoder)

over the past few years deep learning has become a
ties the tiae the fires whae cereiied the seriee s

popular area with deep neural network methods ..
serties thae thae thae seried seaties seaties ..

Ablation Studies (Removing CNNs)



Ablation Studies (Removing CNNs)

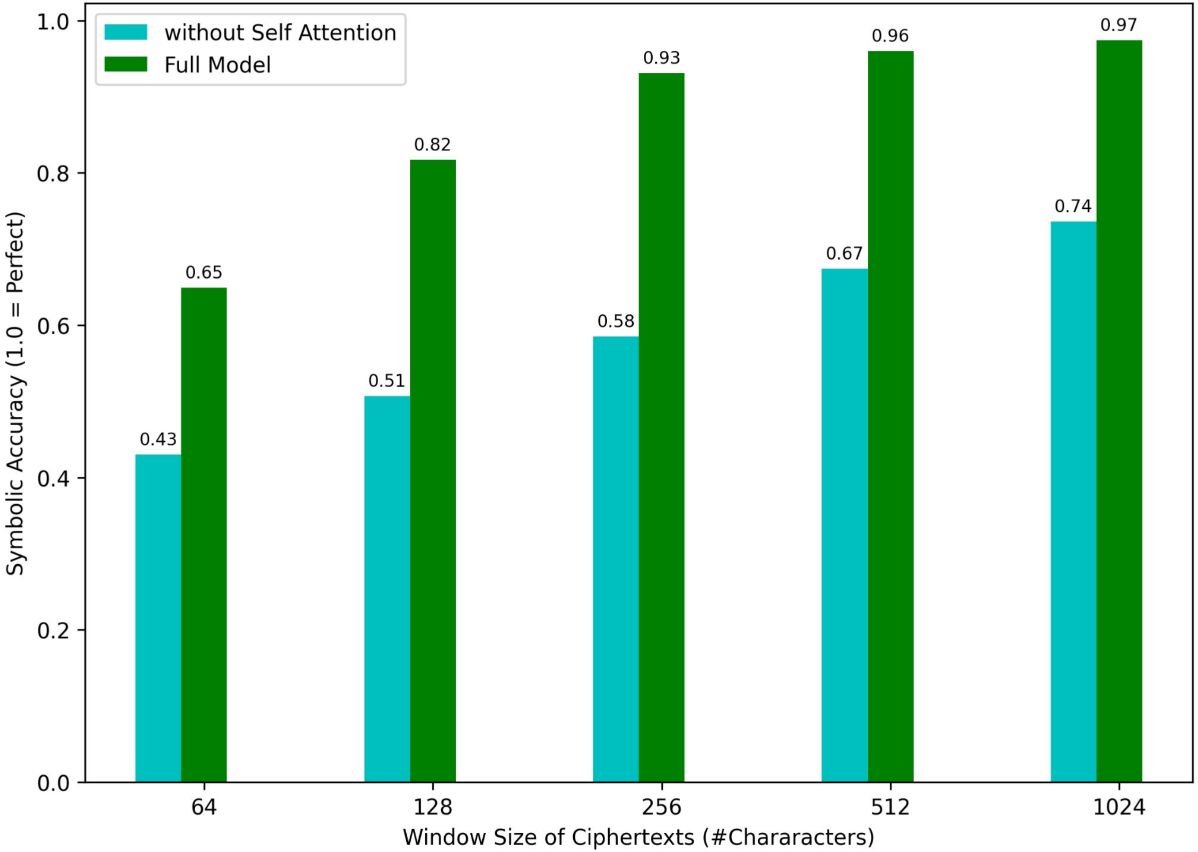
`over the past few years deep learning has become a`
`nyeo tce dtrt feb ketor meed hetotrtd ctr vemnge t`

`popular area with deep neural network methods ..`
`dndmhto toet brtc meed temoth tetbnov getcnmr ..`

Ablation Studies (Removing CNNs)

```
over the past few years deep learning has become a
nyeo tce dtrt feb ketor meed hetotrtd ctr vemnge t

popular area with deep neural network methods ..
dndmhto toet brtc meed temoth tetbnov getcnmr ..
```

Ablation Studies (Removing Self-Attentions)



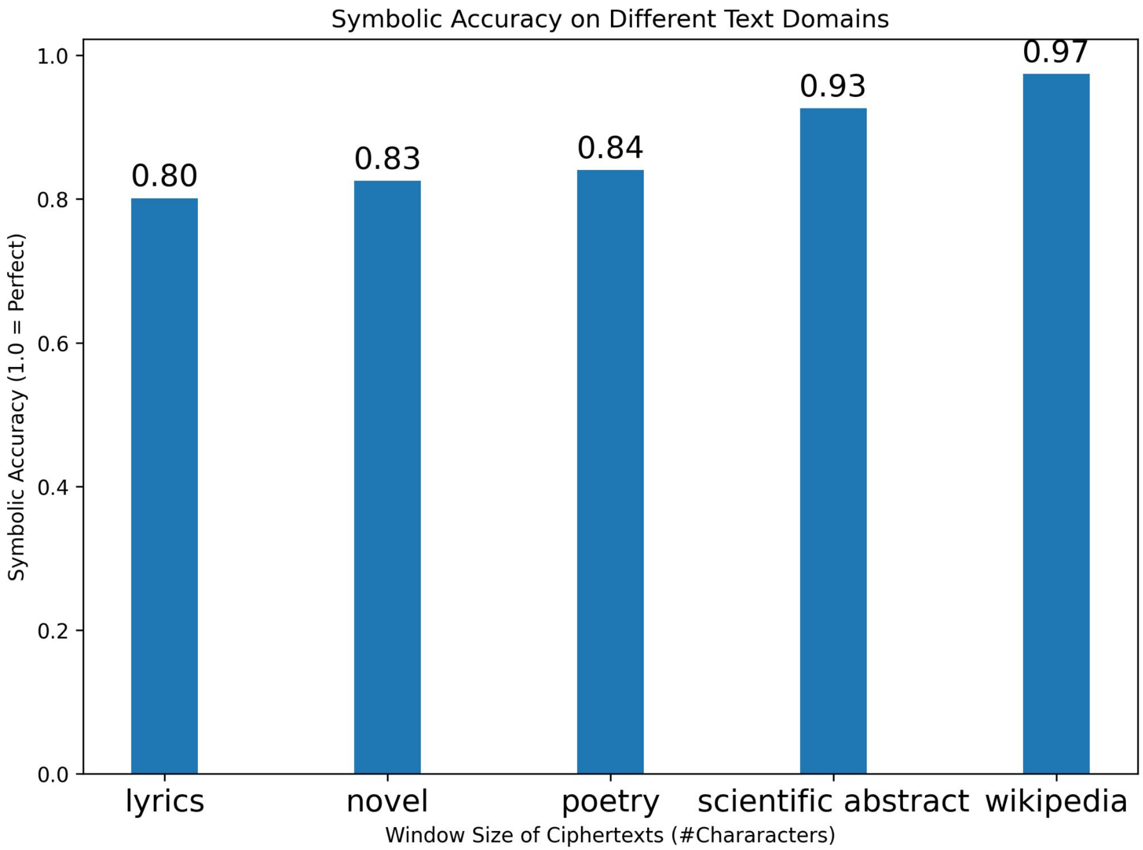
Ablation Studies (Removing Self-Attentions)

```
over the past few years deep learning has become a
oves the mast fey years feen reasonal was became a

popular area with deep neural network methods ..
couulas area from been recond nerwing yerhous ..
```

Ablation Studies (Removing Self-Attentions)

```
over the past few years deep learning has become a
oves the mast fey years feen reasonal was became a

popular area with deep neural network methods ..
couulas area from been recond nerwing yerhous ..
```

Testing Our Model on Various Genres



Performance In Different Genres

Lyrics

is this the real life is this just fantasy caught in a landslide no escape from reality open your eyes look up to the skies see im just a poor boy i need no sympathy because im easy come easy go little high little low any way the wind blows doesnt really matter to me to me

is this the real life is this just fantasy paught in a landslide no espake flom reality open your eyes lonk up to the skies and see im just a ponl boy i need no sympathy bepause im easy pome easy go little high little low any way the king blows doesnt really matter to me to me

Novels

i wasnt sure how i was gonna do it but i knew what i had to do some might ask how someone could have such a strong love for someone they barley even knew the truth is i couldnt even answer that question if i tried everyone

i wasnt sure how i was gonna do it but i knew what i had to do some might ask how someone vould have such a strong love for someone they barley even knew the truth is i vouldnt even answer that question if i tried everyone

Poetry

a little life with dried tubers summer surprised us coming over the starnbergersee with a shower of rain we stopped in the colonnade and went on in sunlight into the hofgarten and drank coffee and talked for an hour bin gar keine russin stamm aus litauen echt deutsch

a mittme mife with dried tupers subber surprised us cobing over the starnpergersee with a shower of rain we stopped in the comonnade and went on in suncight into the hofgarten and drank coffee and tamked for an hour pin gar keine russin stabk aus mitauen echt deutsch

Scientific Article

during character segmentation we test our model on three types of random noise insertion deletion and substitution we experiment with different noise percentages for ciphers of length 256 table 5 we report the results of training and testing on ciphers with only substitution

during character segmentation be test our model on three types of random noise insertion deletion and substitution be experiment with different noise percentages for ciphers of length 100 table 2 be report the results of training and testing on ciphers with only substitution

Future Directions

Looking into cracking substitution ciphers with a larger character space (A-Z, a-z, numbers, other symbols)

Substitution ciphers on the level of sub-words instead of characters

Other cipher systems beyond substitution ciphers

Questions?

References

Yingqiang Gao, Nikola I. Nikolov, Yuhuang Hu, and Richard H.R. Hahnloser. 2020. Character-Level Translation with Self-attention. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1591–1604, Online. Association for Computational Linguistics.

Nada Aldarrab and Jonathan May. 2021. Can Sequence-to-Sequence Models Crack Substitution Ciphers?. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7226–7235, Online. Association for Computational Linguistics.

Cipher Busters, Deep Learning Final Project 2021, post link: https://devpost.com/software/cipherbusters