LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Vision-Based Approach to Noisy Text Recognition

Han Yang & Yian Yu

18.08.2023

Computer Vision and Deep Learning: Automatic Image Understanding and Recognition SoSe23

Guided by Prof. Björn Ommer

Supervisor: Dmytro Kotovenko

Content

- Motivation and Background
- Task Definition and Conceptual Design
- Implementation: Model and Training
- Experiments
  - Dataset
  - Variables and Experimental Condition
- Results
  - Evaluation and Statistical Analysis
  - Case Study
- Conclusion and Future Work

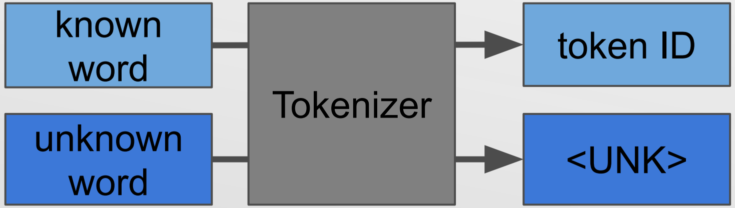
1. Motivation & Background

1. Out-of-Vocabulary (OOV) Problem in Natural Language Processing (NLP)

   unknown words appear in test set but not in training set.

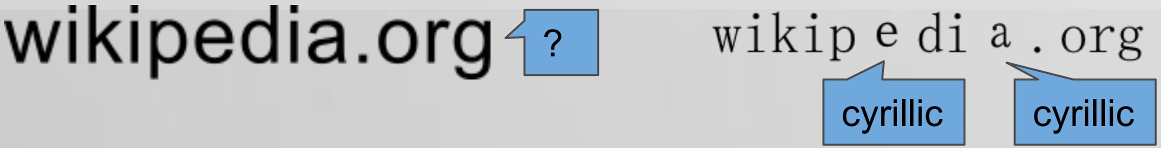
   caused by *small* training set or *noise*

   e.g. “word” → 5, “w0rd” → <UNK>



The OOV problem

2. Human Vision Robustness



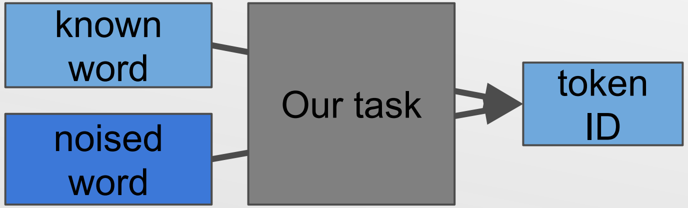
Can we improve **Robustness** again **Noise** by vision-based method?

e.g. “word” → 5, “w0rd” → 5

Image source: https://en.wikipedia.org/wiki/File:IDN_homograph_attack_1.svg

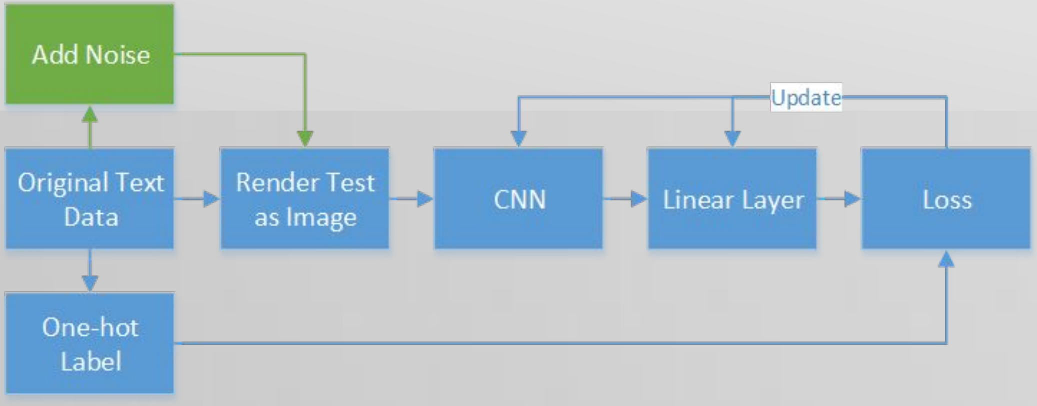
2. Task Definition & Conceptual Design

1. Robustness: OOV problem caused by Noise
2. Our system should recognize noised word and predict a correct Token ID



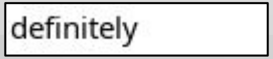
Our task

Methodology: a pipeline



3. Implementation: Model and Training

1. Dictionary:
   a. Word → ID
   b. ID → Word
2. Render Image with Pygame

definitely → 

definitely

3. CNN
   a. 1 Convolutional Layer
   b. Relu
   c. Max Pooling Layer
4. Linear Layer

   Output dimension: Size of Vocabulary
5. Loss: Cross Entropy
6. Optimizer: Adam

Image source: https://www.pygame.org/news

4. Experiments

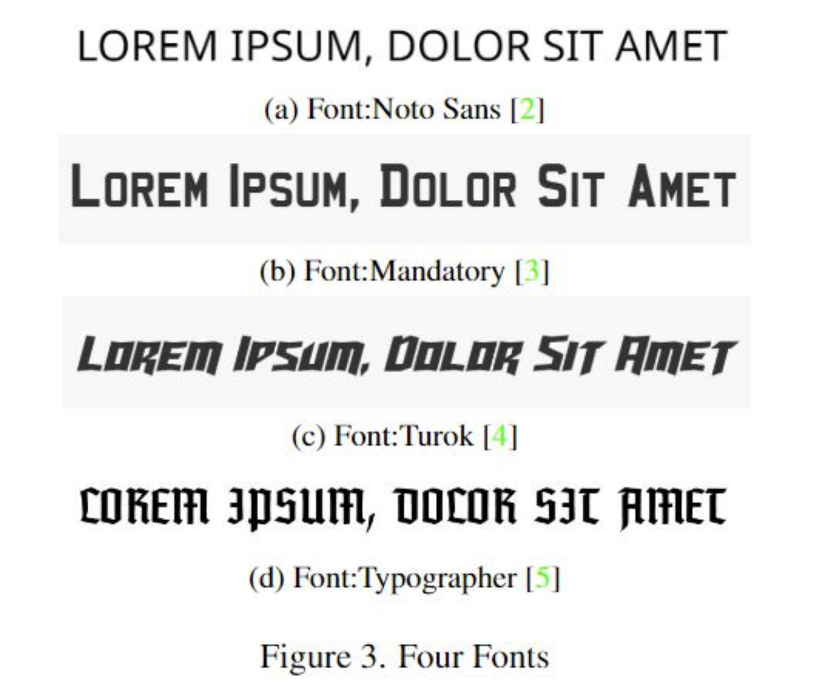
4.1 Dataset

- **Multitarget TED Talks Task (MTTT) Dataset**
  - focus on the **English** portion of the en-de (English-German) translation set
  - count the frequency of each word
  - The most frequent **4571 words** were selected as token

4.2 Variables and Experimental Condition

4 fonts:

- Noto Sans
- Mandatory
- Turok
- Typographer



LOREM IPSUM, DOLOR SIT AMET

(a) Font:Noto Sans [2]

LOREM IPSUM, DOLOR SIT AMET

(b) Font:Mandatory [3]

LOREM IPSUM, DOLOR SIT AMET

(c) Font:Turok [4]

LOREM IPSUM, DOLOR SIT AMET

(d) Font:Typographer [5]

Figure 3. Four Fonts

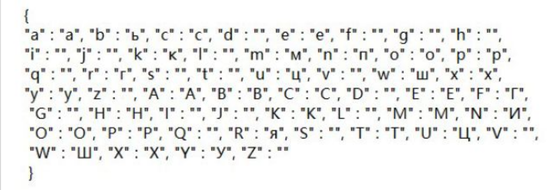
4.2 Variables and Experimental Condition

➔ 3 types of noise: **Greek letters**, **Cyrillic letters**, and **leetspeak**

a → α

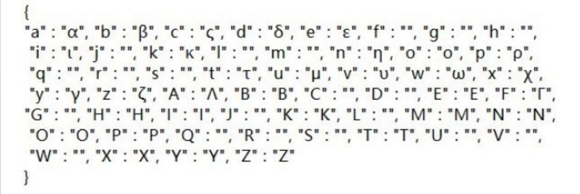
a → a

o → 0



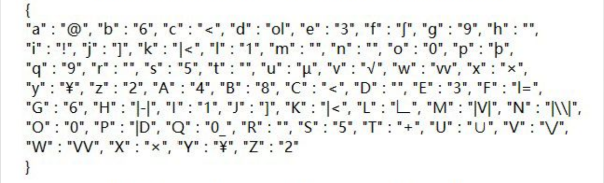
```
{
"a" : "а", "b" : "ь", "c" : "с", "d" : "", "e" : "е", "f" : "", "g" : "", "h" : "",
"i" : "", "j" : "", "k" : "к", "l" : "", "m" : "м", "n" : "п", "o" : "о", "p" : "р",
"q" : "", "r" : "г", "s" : "", "t" : "", "u" : "ц", "v" : "", "w" : "ш", "x" : "х",
"y" : "у", "z" : "", "A" : "А", "B" : "В", "C" : "С", "D" : "", "E" : "Е", "F" : "Г",
 "G" : "", "H" : "Н", "I" : "", "J" : "", "K" : "К", "L" : "", "M" : "М", "N" : "И",
 "O" : "О", "P" : "Р", "Q" : "", "R" : "я", "S" : "", "T" : "Т", "U" : "Ц", "V" : "",
 "W" : "Ш", "X" : "Х", "Y" : "У", "Z" : ""
 }
```

(a) Alphabet to Cyrillic Dictionary Cross Reference



```
{
"a" : "α", "b" : "β", "c" : "ς", "d" : "δ", "e" : "ε", "f" : "", "g" : "", "h" : "",
 "i" : "ι", "j" : "", "k" : "κ", "l" : "", "m" : "", "n" : "η", "o" : "ο", "p" : "ρ",
 "q" : "", "r" : "", "s" : "", "t" : "τ", "u" : "μ", "v" : "υ", "w" : "ω", "x" : "χ",
 "y" : "γ", "z" : "ζ", "A" : "Λ", "B" : "Β", "C" : "", "D" : "", "E" : "Ε", "F" : "Γ",
"G" : "", "H" : "Η", "I" : "Ι", "J" : "", "K" : "Κ", "L" : "", "M" : "Μ", "N" : "Ν",
 "O" : "Ο", "P" : "Ρ", "Q" : "", "R" : "", "S" : "", "T" : "Τ", "U" : "", "V" : "",
 "W" : "", "X" : "Χ", "Y" : "Υ", "Z" : "Ζ"
}
```

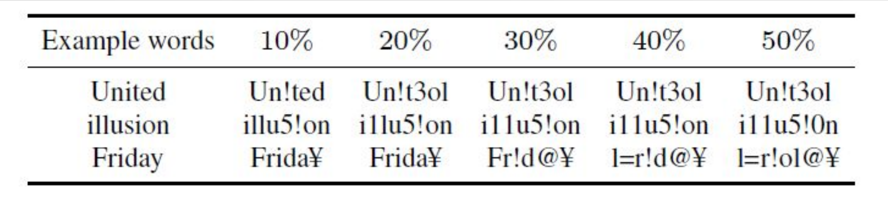
(b) Alphabet to Greek Dictionary Cross Reference



```
{
"a" : "@", "b" : "6", "c" : "<", "d" : "ol", "e" : "3", "f" : "ʃ", "g" : "9", "h" : "",
"i" : "!", "j" : "]", "k" : "|<", "l" : "1", "m" : "", "n" : "", "o" : "0", "p" : "þ",
"q" : "9", "r" : "", "s" : "5", "t" : "", "u" : "µ", "v" : "√", "w" : "vv", "x" : "×",
"y" : "¥", "z" : "2", "A" : "4", "B" : "8", "C" : "<", "D" : "", "E" : "3", "F" : "l=",
"G" : "6", "H" : "|-|", "I" : "1", "J" : "]", "K" : "|<", "L" : "|_", "M" : "|V|", "N" : "|\\|",
"O" : "0", "P" : "|D", "Q" : "0_", "R" : "", "S" : "5", "T" : "+", "U" : "∪", "V" : "\/",
"W" : "VV", "X" : "×", "Y" : "¥", "Z" : "2"
}
```

(c) Alphabet to Leetspeak Dictionary Cross Reference

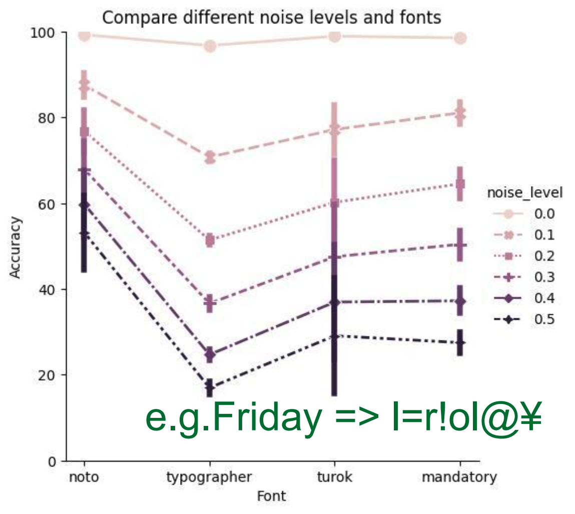
➔ 5 probabilities of **10%**, **20%**, **30%**, **40%** and **50%** for each character replaced



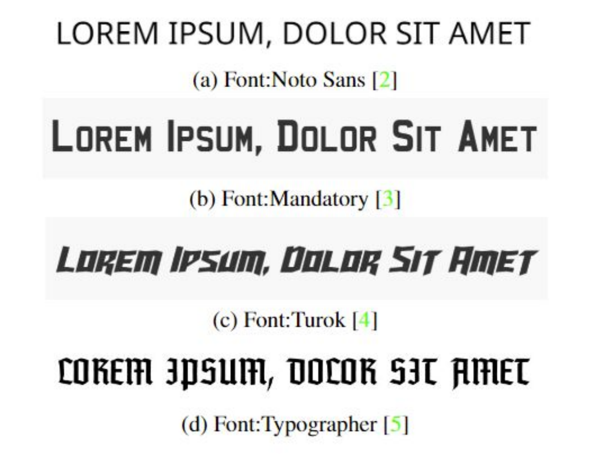
Example words	10%	20%	30%	40%	50%
United	Un!ted	Un!t3ol	Un!t3ol	Un!t3ol	Un!t3ol
illusion	illu5!on	i1lu5!on	i11u5!on	i11u5!on	i11u5!0n
Friday	Frida¥	Frida¥	Fr!d@¥	l=r!d@¥	l=r!ol@¥

5. Results

5.1 Evaluation



e.g.Friday => I=r!ol@¥


LOREM IPSUM, DOLOR SIT AMET
LOREM IPSUM, DOLOR SIT AMET
LOREM IPSUM, DOLOR SIT AMET
LOREM IPSUM, DOLOR SIT AMET

(a) Font:Noto Sans [2]

(b) Font:Mandatory [3]

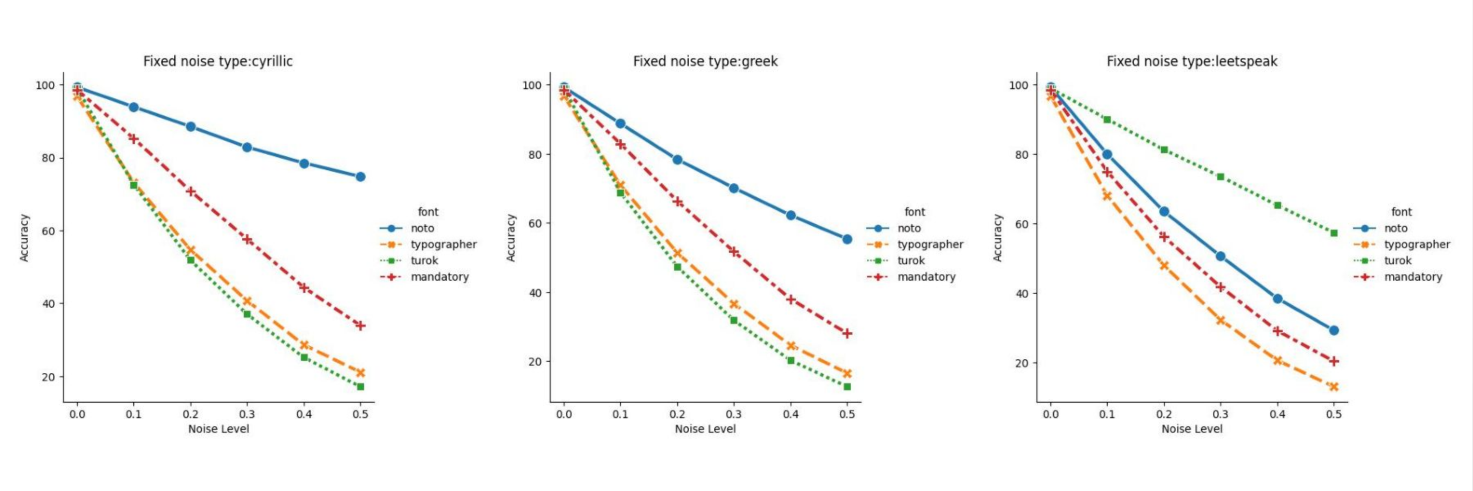
(c) Font:Turok [4]

(d) Font:Typographer [5]

Figure 3. Four Fonts

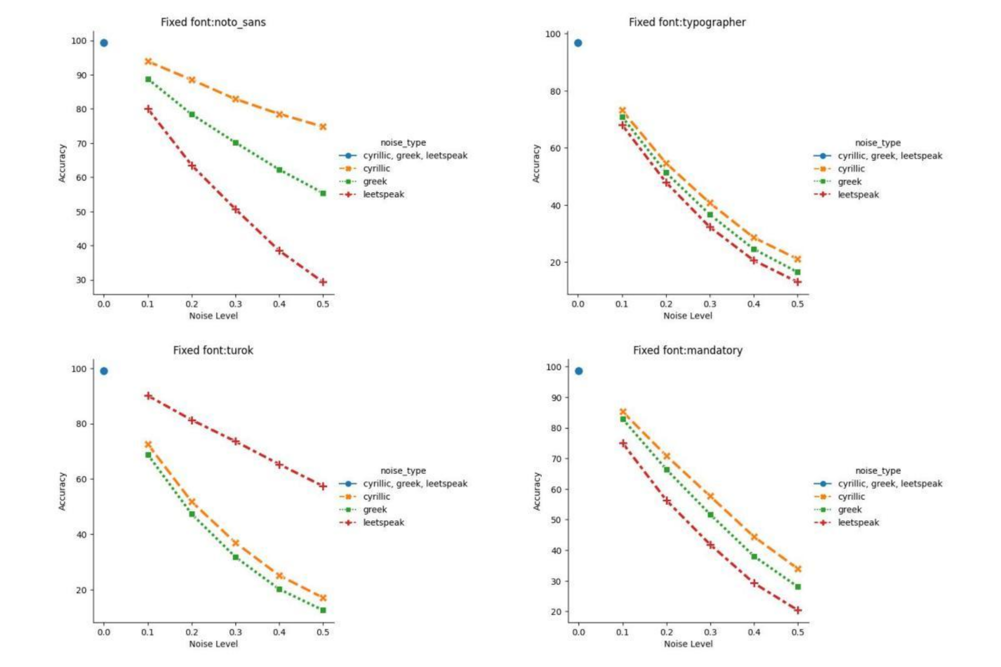
➔ Effect of noise ratio on model robustness in text recognition.

5.1 Evaluation: noise type fixed



➔ Effect of fonts on model robustness in text recognition.

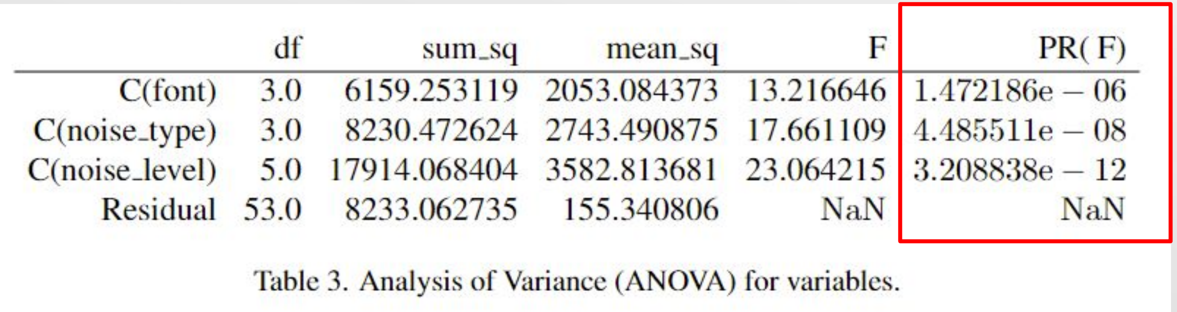
5.1 Evaluation: font fixed



➔ Effect of noise type on model robustness in text recognition.

➔ Challenging: Leetspeak > Greek letters > Cyrillic letters

5.2 Statistical Analysis

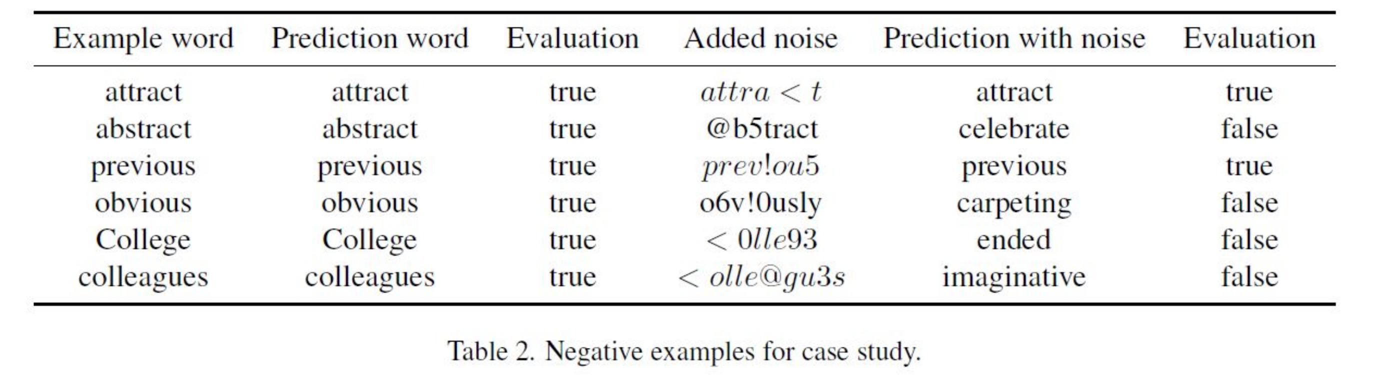


	df	sum_sq	mean_sq	F	PR(F)
C(font)	3.0	6159.253119	2053.084373	13.216646	1.472186e − 06
C(noise_type)	3.0	8230.472624	2743.490875	17.661109	4.485511e − 08
C(noise_level)	5.0	17914.068404	3582.813681	23.064215	3.208838e − 12
Residual	53.0	8233.062735	155.340806	NaN	NaN

Table 3. Analysis of Variance (ANOVA) for variables.

- ➔ All of three variables, namely "Font" "Noise Type" and "Noise Level" have p-values lower than 0.05.
- ➔ Statistical significance: have decisive impact on the model's results for text recognition.

5.3 Case Study



Example word	Prediction word	Evaluation	Added noise	Prediction with noise	Evaluation
attract	attract	true	$attra < t$	attract	true
abstract	abstract	true	@b5tract	celebrate	false
previous	previous	true	$prev!ou5$	previous	true
obvious	obvious	true	o6v!0usly	carpeting	false
College	College	true	$< 0lle93$	ended	false
colleagues	colleagues	true	$< olle@gu3s$	imaginative	false

Table 2. Negative examples for case study.

➔ Result of case study for negative examples:

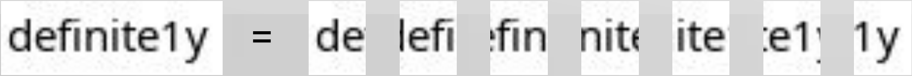
1. No noises added: words with similar characters can be successfully classified.
2. With noises: Classifying words with similar characters becomes challenging.

5. Conclusion and Future Work

- Built a pipeline to improve the Robustness against Noise via Vision Method
- Explored influence of fonts, noise, and noise level
- Case study: robustness against similar words

Limitation & Future Work

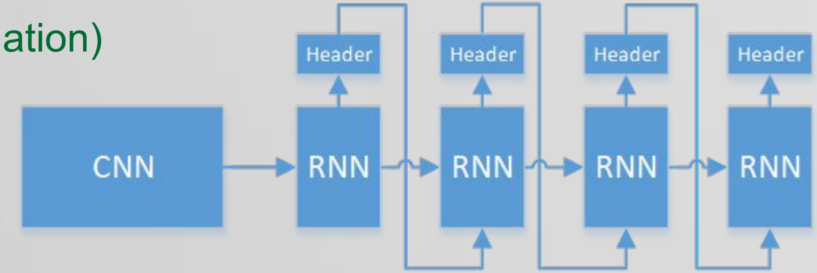
- Since words have **unfixed length**, we could **split** image into **slices**


definite1y
=
de
lefi
efin
nite
ite
e1y
1y

- **Downstream Tasks** (e.g. Machine Translation)

  Now: CNN + Linear Layer (Header)

  Future: CNN + RNN (seq.)



References

[1] https://www.pygame.org/news.

[2] Thomas Bohm. Letter and symbol misrecognition in highly legible typefaces for general, children, dyslexic, visually impaired and ageing readers. Information Design Journal, 21(1):34–50, 2014.

[3] Kevin Duh. The multitarget ted talks task. http:// www.cs.jhu.edu/˜kevinduh/a/multitargettedtalks/, 2018.

[4] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Reading text in the wild with convolutional neural networks. volume 116, pages 1–20. Springer, 2016.

[5] Elizabeth Salesky, David Etter, and Matt Post. Robust openvocabulary translation from visual text representations. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 7235–7252, Online and Punta Cana, Dominican Republic, Nov. 2021. Association for Computational Linguistics. 2,

[6] Baoguang Shi, Xiang Bai, and Cong Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE transactions on pattern analysis and machine intelligence, 39(11):2298–2304, 2016.

Thank you for listening

Han Yang & Yian Yu
18.08.2023