An End-to-End Approach for Handwriting Recognition: From Handwritten Text Lines to Complete Pages

Dayvid Castro
Universidade Federal de Pernambuco

Byron Bezerra
Universidade de Pernambuco

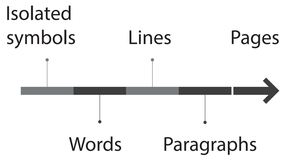
Cleber Zanchettin
Universidade Federal de Pernambuco

1. Introduction

What is offline Handwritten Text Recognition (HTR)?

Handwritten text image → Machine-Encoded text

Advancements in Handwritten Text Recognition



Handwritten Document Recognition (HDR):
Text + Layout Tokens

Handwritten Document Recognition (HTR) Task

Given document image **D**

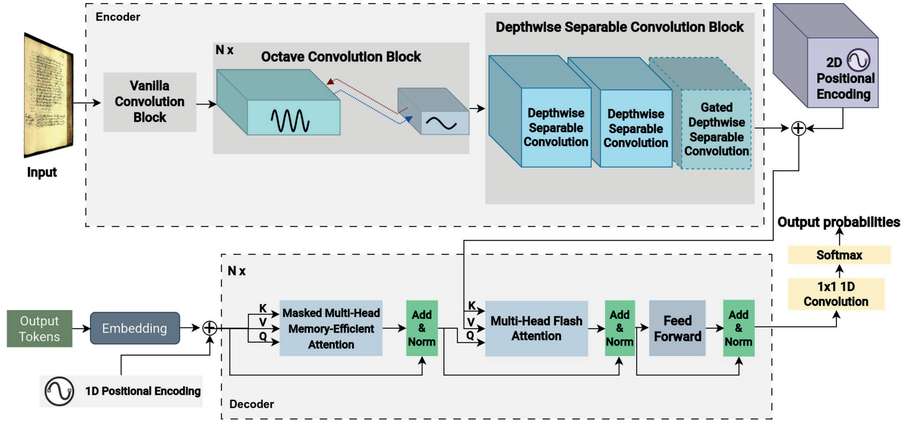
Find the most likely token sequence $t^* = (t_1 \ldots t_n)_{t \in T}$

Where $T = A \cup L \cup \{<eot>\}$

- A → Alphabet tokens
- L → Layout tokens
- $\{<eot>\}$ → End-of-transcription token

2. Method

Document Attention Network for Computationally Efficient Recognition (DANCER)



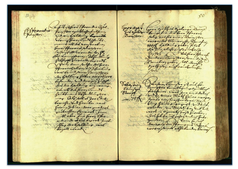
3. Experiments

Data:
ICFHR READ 2016

1- Single-Page Task
2- Double-Page Task

1

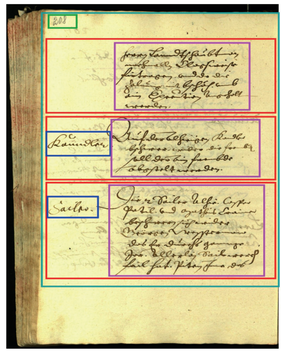
2



Comparison with the State of the Art

Method	Single-Page				Double-Page			
	CER↓ (%)	WER↓ (%)	LOER↓ (%)	mAP_{CER}↑ (%)	CER↓ (%)	WER↓ (%)	LOER↓ (%)	mAP_{CER}↑ (%)
DAN	3.43	**13.05**	5.17	93.32	3.70	14.15	4.98	93.09
Faster DAN	3.95	14.06	3.82	94.20	3.88	14.97	**3.08**	94.54
DANCER	**3.36**	13.73	**3.37**	**94.73**	**3.37**	**13.34**	3.91	**95.08**

DANCER Prediction Example

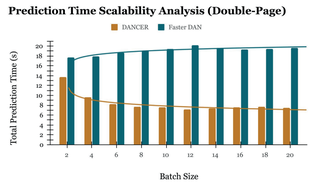


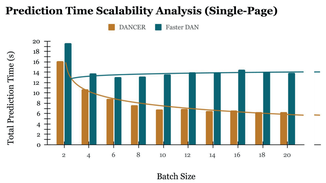
Ground Truth

```
<page>
 <number>208</number>
 <section>
  <body>herrn Lanndtshaübtman
    nochmalin Clagsweise
    firtragen. vnd da die
    Zalüng nit bschöh. vmb
    ain Execütion Angehalt
    werden.
  </body>
 </section>
 <section>
  <annotation>Kandler.</annotation>
  <body>Aüf der Alheigen Kandler
    bschwer. wider die frembd
    soll der Ain frembde
    Abgstelt werden.
  </body>
 </section>
 <section>
  <annotation>Sailer.</annotation>
  <body>Die .2 Sailer Alhie. Caspar
    Patil. vnd Matheis Crainer
    beschwern sich wider
    Geörgen Wesstermair
    das Er dürchs gannze
    Jar. Allerlaj Sail werch
    fail hat. Piten Ime das
  </body>
 </section>
</page>
```

Prediction

```
<page>
 <number>208</number>
 <section>
  <body>herrn Lanndtshaübtman
    nochmalin Clagsweise
    firtragen. vnd da die
    Zalüng nit bschöh. vmb
    ain Execütion Angehalt
    werden,
  </body>
 </section>
 <section>
  <annotation>Kanndler.</annotation>
  <body>Aüf der Alheigen Kandler
    bschwer. wider die frembd
    soll der Ain frembde
    Abgstelt werden.
  </body>
 </section>
 <section>
  <annotation>Sailer.</annotation>
  <body>Die .2 Sailer Alhie. Caspar
    Patil. vnd Matheis Crainer
    beschwern sich wider
    Geörgen Wesstermair
    das Er dürchs gannze
    Jar. Allerlai Sail werch
    fail hat. Piten Ime das
  </body>
 </section>
</page>
```





4. Contributions

New HDR model with key innovations over previous works:

- Use of a gated depthwise separable convolution to induce spatial attention.
- Use of octave convolutions to reduce spatial redundancy and improve performance.
- Integration of Memory-Efficient and Flash Attention operations

Comprehensive scalability analyses reveal DANCER's advantages as workload increases.