

# Finetune your OCR!

## Improving automated text recognition for early printed works by finetuning existing Tesseract models

fonen mit 2 337 812 074 *M* versichert blieben, davon 552 246  
Personen mit 1 847 622 742 *M* bei den 35 Se im  
Deutschen Reich. Der gesammte Zuwachs im Jahre 1877 stellt sich  
auf 15 777 Personen mit 110 873 *M*, d. h. 2,14 % der  
Zahl der Versicherten und +,98 % der wachen,  
summe. Die Berücksichtigung der Unfälle, welche die  
Liquidation einiger Gesellschafenschaften im Jahre 1877 zur Folge gehabt  
hat vermindert sich der Zuwachs auf 8542 (1,16%) der Versicherten

Jan Kamlah / Thomas Schmidt

# Finetune your OCR! Improving automated text recognition for early printed works by finetuning existing Tesseract models



1. OCR and work-specific training at UB Mannheim
2. Introduction
3. Training with synthetic ground truth (Approach A)
4. Training with real-world ground truth (Approach B)
5. Conclusion

# Finetune your OCR! Improving automated text recognition for early printed works by finetuning existing Tesseract models



1. OCR and work-specific training at UB Mannheim
2. Introduction
3. Training with synthetic ground truth (Approach A)
4. Training with real-world ground truth (Approach B)
5. Conclusion

# 1. OCR and work-specific training at UB Mannheim

- since 2022: work-specific training (for eScriptorium + Kraken) with theological prints, 1860-72 (DigiTheo Ground Truth: <https://github.com/UB-Mannheim/DTGT>)
- since 2021: module project „work-specific training“ (for Tesseract + Calamari) as part of the DFG-funded initiative „OCR-D“
- 2013–2019: digitisation / OCR „Aktienführer-Datenarchiv“ (German stock exchange data 1979-99)
- 2019: work-specific training for „Weisthümer“ (Jacob Grimm, 1840) <https://github.com/UB-Mannheim/Weisthuemer>
- 2011–2018: digitisation / OCR „Desbillons“-collection (French clergyman and scholar)



# Finetune your OCR! Improving automated text recognition for early printed works by finetuning existing Tesseract models



1. OCR and work-specific training at UB Mannheim
2. **Introduction**
3. Training with synthetic ground truth (Approach A)
4. Training with real-world ground truth (Approach B)
5. Conclusion

## 2. Introduction

Im Ganzen waren im Jahre 1877 815272 Personen mit 2511478651 *M.* versichert; hiervon schieden 13109 Personen (35906127 *M.*) durch Tod und 48754 Personen (137760445 *M.*) anderweitig aus, so daß am Jahresluß 753409 Personen mit 2337812074 *M.* versichert blieben, davon 552246 Personen mit 1847622742 *M.* bei den 35 Gesellschaften im Deutschen Reich. Der gesammte Zuwachs im Jahre 1877 stellt sich auf 15777 Personen mit 110873820 *M.*, d. h. 2,14% der Zahl der Versicherten und 4,98% der Versicherungssumme. Die Berücksichtigung der Unfälle, welche die Liquidation einiger Gesellschaften im Jahre 1877 zur Folge gehabt hat, vermindert sich der Zuwachs auf 8542 (1,16%) der Versicherten und 93396677 *M.* (4,19%) der Summe (gegen 4,35 bzw. 6,57% in 1876; 5,30 bzw. 8,09% in 1875; 7,12 bzw. 9,29% in 1874; 8,84 bzw. 11,44% in 1873). Der Durchschnitt der Summe belief sich pro Kopf im Jahre 1877 auf 3103 *M.*, gegen 3021 *M.* in 1876. Den stärksten Zuwachs hatten im Jahre 1877 von den 35 Gesellschaften im Deutschen Reich diejenigen zu Karlsruhe (2810616 *M.*) und Gotha (1167800 *M.*), außerdem hatten noch 4 dieser Gesellschaften einen Zuwachs, alle übrigen eine Abnahme der Versicherungssumme erlitten.

*Deutscher Reichsanzeiger und Preußischer  
Staatsanzeiger, 1878, Issue 248*



## 2. Introduction

Im Ganzen waren im Jahre 1877 815272 Personen mit 2 511 478 651 *M.* versichert; hiervon schieden 13 109 Personen (35 906 127 *M.*) durch Tod und 48 754 Personen (137 760 445 *M.*) anderweitig aus, so daß am Jahreschluß 753 409 Personen mit 2 337 812 074 *M.* versichert blieben, davon 552 246 Personen mit 1 847 622 742 *M.* bei den 35 Gesellschaften im Deutschen Reich. Der gesammte Zuwachs im Jahre 1877 stellt sich auf 15 777 Personen mit 110 873 820 *M.*, d. h. 2,14 % der Zahl der Versicherten und 4,98 % der Versicherungssumme. Die Berücksichtigung der Unfälle, welche die Liquidation einiger Gesellschaften im Jahre 1877 zur Folge gehabt hat, vermindert sich der Zuwachs auf 8542 (1,16 %) der Versicherten und 93 396 677 *M.* (4,19 %) der Summe (gegen 4,35 bzw. 6,57 % in 1876; 5,30 bzw. 8,09 % in 1875; 7,12 bzw. 9,29 % in 1874; 8,84 bzw. 11,44 % in 1873). Der Durchschnitt der Summe belief sich pro Kopf im Jahre 1877 auf 3103 *M.*, gegen 3021 *M.* in 1876. Den stärksten Zuwachs hatten im Jahre 1877 von den 35 Gesellschaften im Deutschen Reich diejenigen zu Karlsruhe (2 810 616 *M.*) und Gotha (1 167 800 *M.*), außerdem hatten noch 4 dieser Gesellschaften einen Zuwachs, alle übrigen eine Abnahme der Versicherungssumme erlitten.

*Deutscher Reichsanzeiger und Preußischer Staatsanzeiger*, 1878, Issue 248

Im Ganzen waren im Jahre 1877 815272 Personen mit 2511478 651 -ÄZ erfichert; hiervon schieden 13 109 Personen (35 906 127 ii) durch Tod und 48 754 11 (137 760 445 kii) anderweitig aus, so daß am Jahreschluß 753 409 Per= fonen mit 2337 812 074 &a versichert blieben, davon 552 246 Personen mit 1847 622 742 bei den 35 re 157 fteh im Deutschen Reich. Der gesammte Zuwachs im Jahre 1877 ftellt sich auf 15 777 Perionen mit 110 873 820 , d. h. 2,14 % der Zahl der Versicherten und 4,98 % der Wunche die Summe. Die Berücksichtigung der Unfälle, welche die Liquidation einiger Gesellschaften im Jahre 1877 zur Folge gehabt hat, vermindert sich der Zuwachs auf 8542 (1,16%) der Versicherten und 93 396 677 α (4,19%) der Summe (gegen 4,35 bzw. 6,57 / in 1876; 5,30 bzw. 8,09 % in 1875; 7,12 bzw. 9,29 % in 1874; 8,84 bzw. 11,44% in 187, Der Durchschnitt der Summe belief sich pro Kopf im Jahre 1877 auf 3103 α, gegen 302 σ in 1876. Den stärksten Zuwachs hatten im Jahre 1877 von den 35 Gefell= schaften im Deutschen Reich diejenigen zu Karlsruhe (2810 616 und Gotha (1 167 800 ), außerdem hatten noch 4 dieser Gefell= schaften einen Zuwachs, alle übrigen eine Abnahme der Verfiche= rungssumme erlitten.

OCR result  
Tesseract (tesseract 5.1.0-20)  
Model: **frak2021 (UB Mannheim)**

## 2. Introduction

Im Ganzen waren im Jahre 1877 815272 Personen mit 2 511 478 651 *M.* versichert; hiervon schieden 13 109 Personen (35 906 127 *M.*) durch Tod und 48 754 Personen (137 760 445 *M.*) anderweitig aus, so daß am Jahreschluß 753 409 Personen mit 2 337 812 074 *M.* versichert blieben, davon 552 246 Personen mit 1 847 622 742 *M.* bei den 35 Gesellschaften im Deutschen Reich. Der gesammte Zuwachs im Jahre 1877 stellt sich auf 15 777 Personen mit 110 873 820 *M.*, d. h. 2,14 % der Zahl der Versicherten und 4,98 % der Versicherungssumme. Die Berücksichtigung der Unfälle, welche die Liquidation einiger Gesellschaften im Jahre 1877 zur Folge gehabt hat, vermindert sich der Zuwachs auf 8542 (1,16 %) der Versicherten und 93 396 677 *M.* (4,19 %) der Summe (gegen 4,35 bzw. 6,57 % in 1876; 5,30 bzw. 8,09 % in 1875; 7,12 bzw. 9,29 % in 1874; 8,84 bzw. 11,44 % in 1873). Der Durchschnitt der Summe belief sich pro Kopf im Jahre 1877 auf 3103 *M.* gegen 3021 *M.* in 1876. Den stärksten Zuwachs hatten im Jahre 1877 von den 35 Gesellschaften im Deutschen Reich diejenigen zu Karlsruhe (2 810 616 *M.*) und Gotha (1 167 800 *M.*), außerdem hatten noch 4 dieser Gesellschaften einen Zuwachs, alle übrigen eine Abnahme der Versicherungssumme erlitten.

*Deutscher Reichsanzeiger und Preußischer Staatsanzeiger*, 1878, Issue 248

Im Ganzen waren im Jahre 1877 815272 Personen mit 2511478 651 *-ÄZ* erfichert; hiervon schieden 13 109 Personen (35 906 127 *ii*) durch Tod und 48 754 11 (137 760 445 *kli*) anderweitig aus, so daß am Jahreschluß 753 409 Personen mit 2337 812 074 *&* versichert blieben, davon 552 246 Personen mit 1847 622 742 bei den 35 re 157 fteh im Deutschen Reich. Der gesammte Zuwachs im Jahre 1877 ftellt sich auf 15 777 Perionen mit 110 873 820 , d. h. 2,14 % der Zahl der Versicherten und 4,98 % der Wunche die Summe. Die Berücksichtigung der Unfälle, welche die Liquidation einiger Gesellschaften im Jahre 1877 zur Folge gehabt hat, vermindert sich der Zuwachs auf 8542 (1,16%) der Versicherten und 93 396 677 *α* (4,19%) der Summe (gegen 4,35 bzw. 6,57 / in 1876; 5,30 bzw. 8,09 % in 1875; 7,12 bzw. 9,29 % in 1874; 8,84 bzw. 11,44% in 187, Der Durchschnitt der Summe belief sich pro Kopf im Jahre 1877 auf 3103 *α*, gegen 3021 *σ* in 1876. Den stärksten Zuwachs hatten im Jahre 1877 von den 35 Gesellschaften im Deutschen Reich diejenigen zu Karlsruhe (2810 616 *σ*) und Gotha (1 167 800 ), außerdem hatten noch 4 dieser Gesellschaften einen Zuwachs, alle übrigen eine Abnahme der Versicherungssumme erlitten.

OCR result

Tesseract (tesseract 5.1.0-20)

Model: *frak2021* (UB Mannheim)



## 2. Introduction

Im Ganzen waren im Jahre 1877 815272 Personen mit 2 511 478 651 *M.* versichert; hiervon schieden 13 109 Personen (35 906 127 *M.*) durch Tod und 48 754 Personen (137 760 445 *M.*) anderweitig aus, so daß am Jahreschluß 753 409 Personen mit 2 337 812 074 *M.* versichert blieben, davon 552 246 Personen mit 1 847 622 742 *M.* bei den 35 Gesellschaften im Deutschen Reich. Der gesammte Zuwachs im Jahre 1877 stellt sich auf 15 777 Personen mit 110 873 820 *M.*, d. h. 2,14 % der Zahl der Versicherten und 4,98 % der Versicherungssumme. Die Berücksichtigung der Unfälle, welche die Liquidation einiger Gesellschaften im Jahre 1877 zur Folge gehabt hat, vermindert sich der Zuwachs auf 8542 (1,16 %) der Versicherten und 93 396 677 *M.* (4,19 %) der Summe (gegen 4,35 bzw. 6,57 % in 1876; 5,30 bzw. 8,09 % in 1875; 7,12 bzw. 9,29 % in 1874; 8,84 bzw. 11,44 % in 1873). Der Durchschnitt der Summe belief sich pro Kopf im Jahre 1877 auf 3103 *M.* gegen 3021 *M.* in 1876. Den stärksten Zuwachs hatten im Jahre 1877 von den 35 Gesellschaften im Deutschen Reich diejenigen zu Karlsruhe (2 810 616 *M.*) und Gotha (1 167 800 *M.*), außerdem hatten noch 4 dieser Gesellschaften einen Zuwachs, alle übrigen eine Abnahme der Versicherungssumme erlitten.

*Deutscher Reichsanzeiger und Preußischer Staatsanzeiger*, 1878, Issue 248

Im Ganzen waren im Jahre 1877 815272 Personen mit 2511 478 651 *M.* versichert; hiervon schieden 13 109 Personen (35 906 127 *M.*) durch Tod und 48 754 14. (137 760 445 *M.*) anderweitig aus, so daß am Jahreschluß 753 409 Personen mit 2 337 812 074 *M.* versichert blieben, davon 552 246 Personen mit 1 847 622 742 *M.* bei den 35 Se im Deutschen Reich. Der gesammte Zuwachs im Jahre 1877 stellt sich auf 15777 Personen mit 110 873 820 *M.*, d. h. 2,14 % der Zahl der Versicherten und 4,98 % der Wachen, summe. Die Berücksichtigung der Unfälle, welche die Liquidation einiger Gesellschaften im Jahre 1877 zur Folge gehabt hat, vermindert sich der Zuwachs auf 8542 (1,16%) der Versicherten und 93 396 677 *M.* (4,19 %) der Summe (gegen 4,35 bzw. 6,57 % in 1876; 5,30 bzw. 8,09% in 1875; 7,12 bzw. 9,29 in 1874; 8,84 bzw. 11,44 % in 71. Der Durchschnitt der Summe belief sich pro Kopf im Jahre 1877 auf 3103 *M.*, gegen 3021 *M.* in 1876. Den stärksten Zuwachs hatten im Jahre 1877 von den 35 Gesellschaften im Deutschen Reich diejenigen zu Karlsruher (2 810 616 *M.*) und Gotha (1 167 800 *M.*), außerdem hatten noch 4 dieser Gesellschaften einen Zuwachs, alle übrigen eine Abnahme der Versicherungssumme erlitten.

OCR result

Tesseract (tesseract 5.1.0-20)

Model: [frak2021 \(UB Mannheim\)](#) [after work-specific training](#)

## 2. Introduction



Reasons for a work-specific training ...

## 2. Introduction



Reasons for a work-specific training ...

- improve text recognition quality

## 2. Introduction

Reasons for a work-specific training ...

- improve text recognition quality
- expand the character set to be recognized → historical currency symbols, astronomical or mathematical signs, etc.

## 2. Introduction

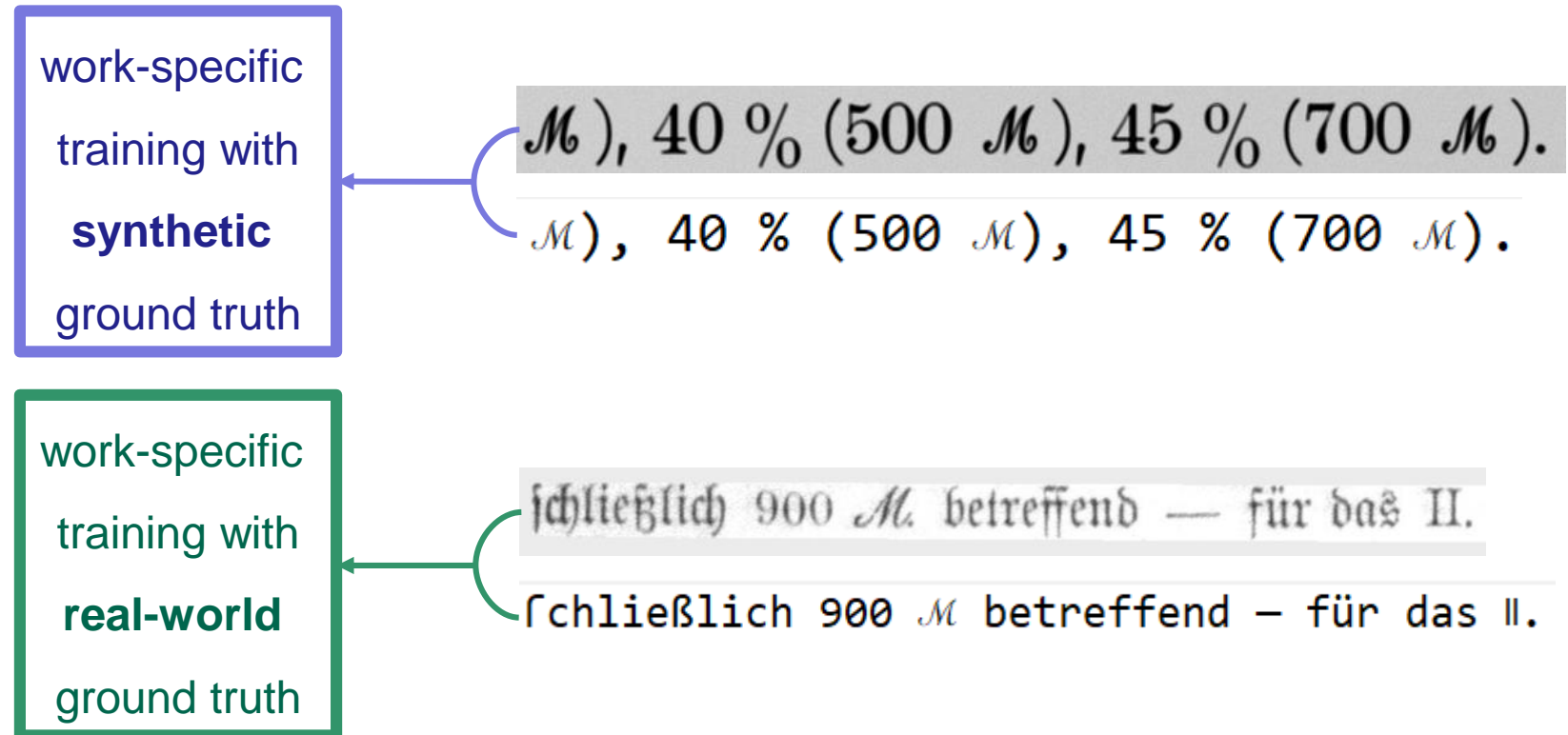
Reasons for a work-specific training ...

- improve text recognition quality
- expand the character set to be recognized → historical currency symbols, astronomical or mathematical signs, etc.
- specific mappings → for example: normalizations, resolution of abbreviations, dealing with white spaces (letterspaced text), etc.



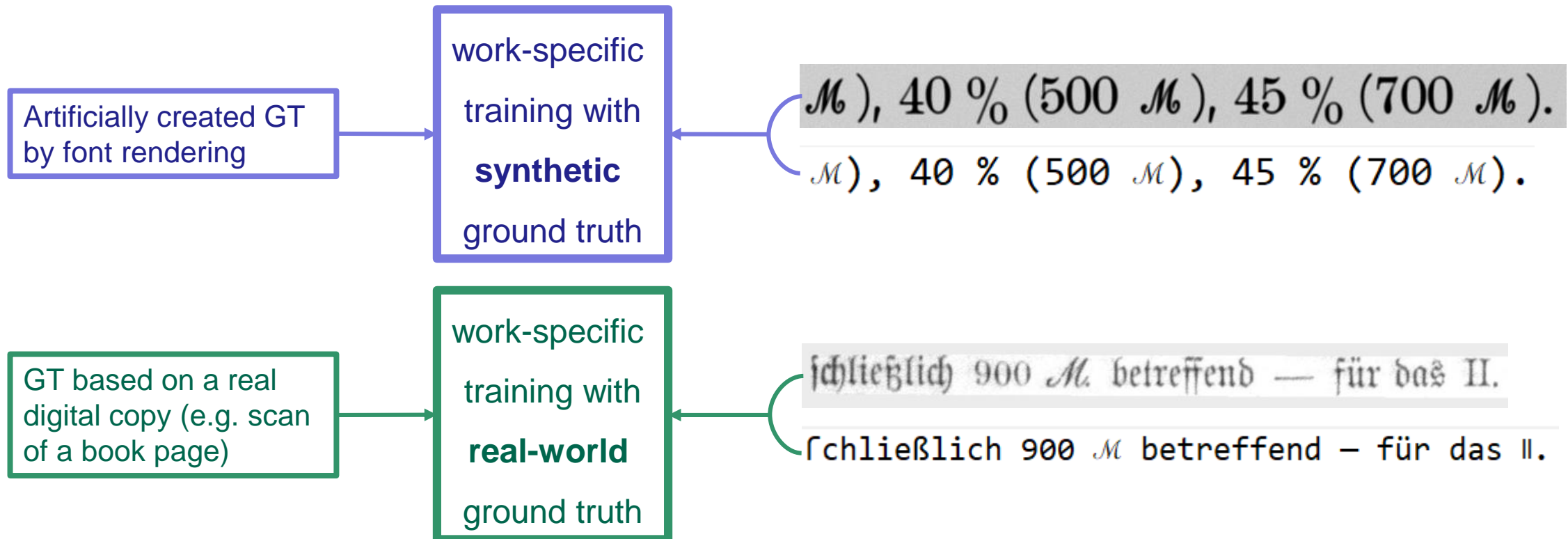
## 2. Introduction

Two tested approaches:



## 2. Introduction

Two tested approaches:



# Finetune your OCR! Improving automated text recognition for early printed works by finetuning existing Tesseract models



1. OCR and work-specific training at UB Mannheim
2. Introduction
3. **Training with synthetic ground truth (Approach A)**
4. Training with real-world ground truth (Approach B)
5. Conclusion

### 3. Training with synthetic ground truth (Approach A)

- **Text2Image** (Tesseract tool) renders text lines from text source and font

### 3. Training with synthetic ground truth (Approach A)

- **Text2Image** (Tesseract tool) renders text lines from text source and font

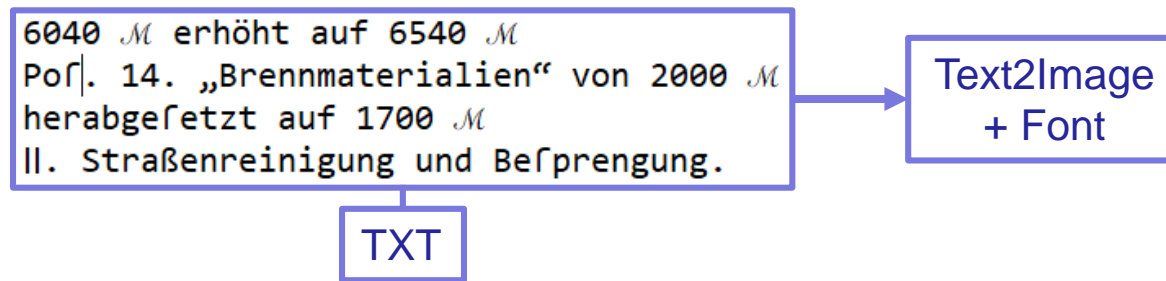
6040 *ℳ* erhöht auf 6540 *ℳ*  
Poſ|. 14. „Brennmaterialien“ von 2000 *ℳ*  
herabgeſetzt auf 1700 *ℳ*  
II. Straßenreinigung und Befprengung.

TXT



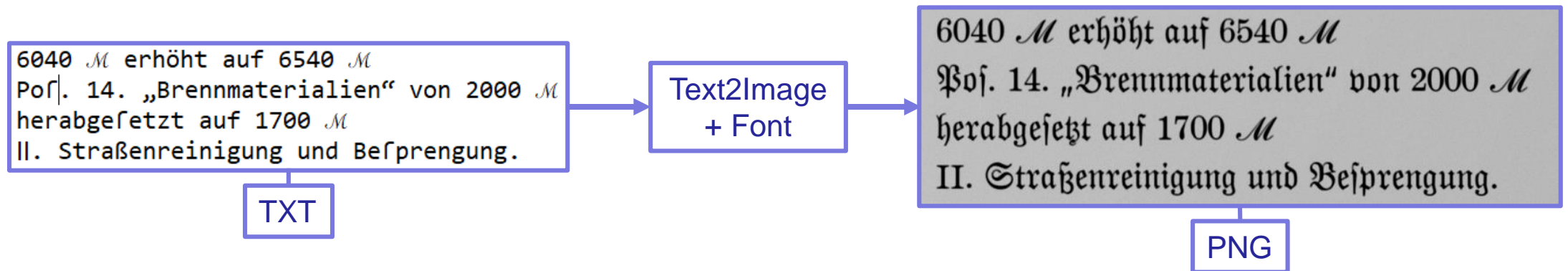
### 3. Training with synthetic ground truth (Approach A)

- **Text2Image** (Tesseract tool) renders text lines from text source and font



### 3. Training with synthetic ground truth (Approach A)

- **Text2Image** (Tesseract tool) renders text lines from text source and font



### 3. Training with synthetic ground truth (Approach A)

- 2 fonts used for rendering:

Gründerfrach=Fraktur	Berthold Mainzer Fraktur
<b>corresponds</b> to the font used in the source material	<b>resembles</b> the font used in the source material
large character set: long f, round r, M etc.	large character set: long f, round r, M etc.

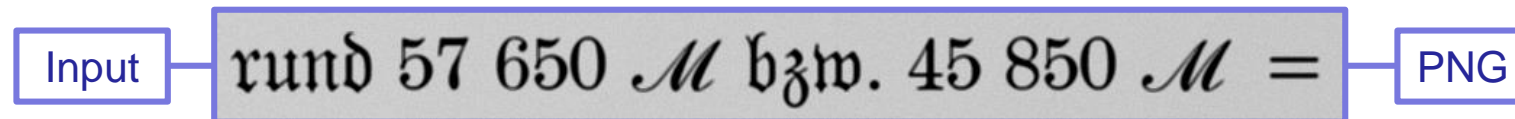
*Table 1: Fonts used for rendering*

### 3. Training with synthetic ground truth (Approach A)

- Additional script to enrich 50 % of the image lines per data augmentation

### 3. Training with synthetic ground truth (Approach A)

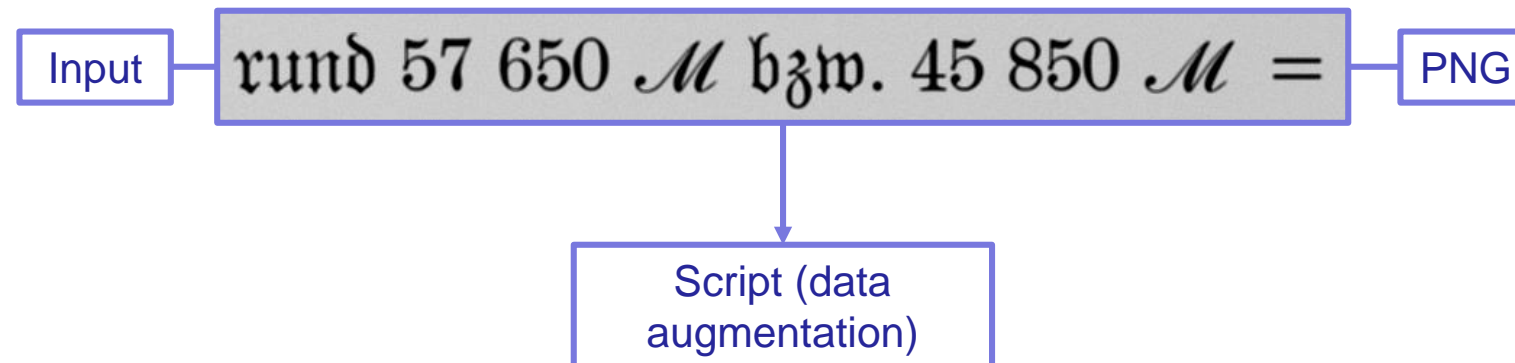
- Additional script to enrich 50 % of the image lines per data augmentation





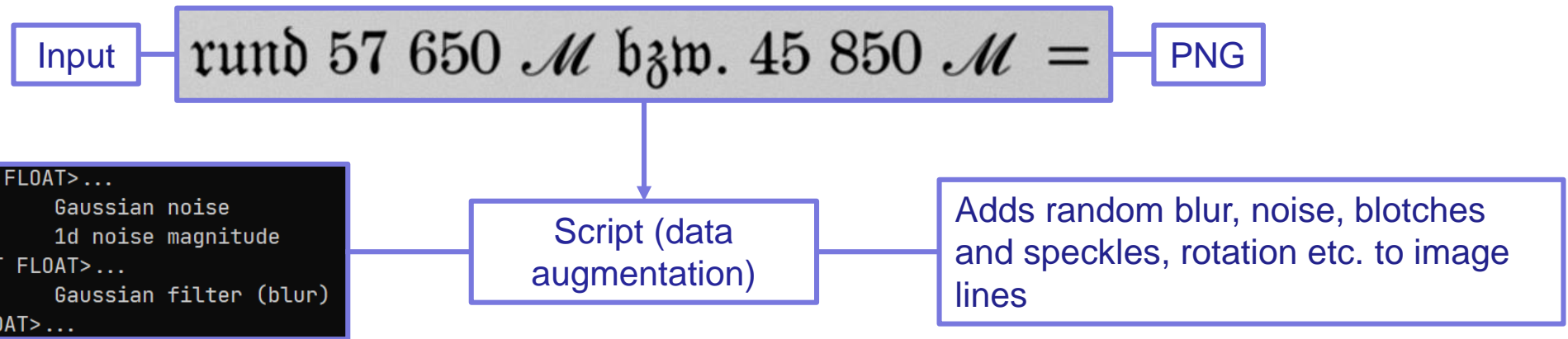
### 3. Training with synthetic ground truth (Approach A)

- Additional script to enrich 50 % of the image lines per data augmentation



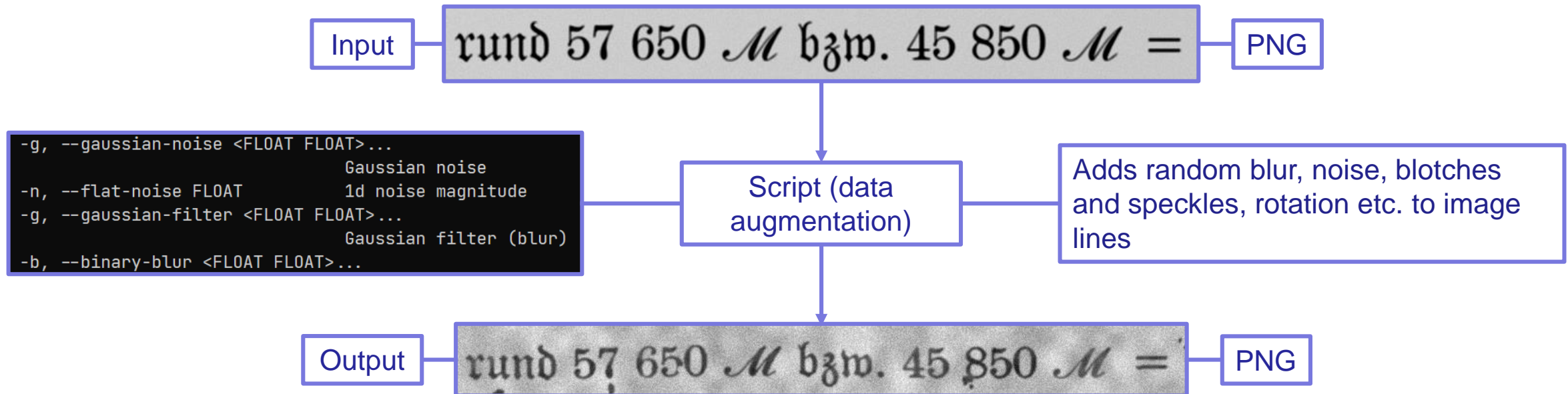
### 3. Training with synthetic ground truth (Approach A)

- Additional script to enrich 50 % of the image lines per data augmentation





### 3. Training with synthetic ground truth (Approach A)

- Additional script to enrich 50 % of the image lines per data augmentation



### 3. Training with synthetic ground truth (Approach A)

Table 2: Training sets with synthetic ground truth

Work-specific training with synthetic ground truth (2 Fonts / 6 Datasets)			
Base model:	frak2021 (UB Mannheim)		
GT source:	„Charlottenburger Amtsschrifttum“, 1879–1919 (ZLB)		
Fonts:	 © 2021 Philipp Poll	 © Peter Wiegel	
Synthetic GT:	<b>LARGE*</b> 3.000 GT lines	<b>MEDIUM†</b> 1.000 GT lines	<b>SMALL†</b> 100 GT lines
	PNG, grayscale, 50 % enriched by data augmentation		
Target:	Extension of the character set: $\mathcal{M}$		
Runtime:	60 epochs with termination criterion → Actual runtime approx. 40–50 epochs		

\* set contains lines with and without symbol  $\mathcal{M}$ ; † set contains only lines with symbol  $\mathcal{M}$

### 3. Training with synthetic ground truth (Approach A)

Evaluation: *Gründerkrach=Fraktur*

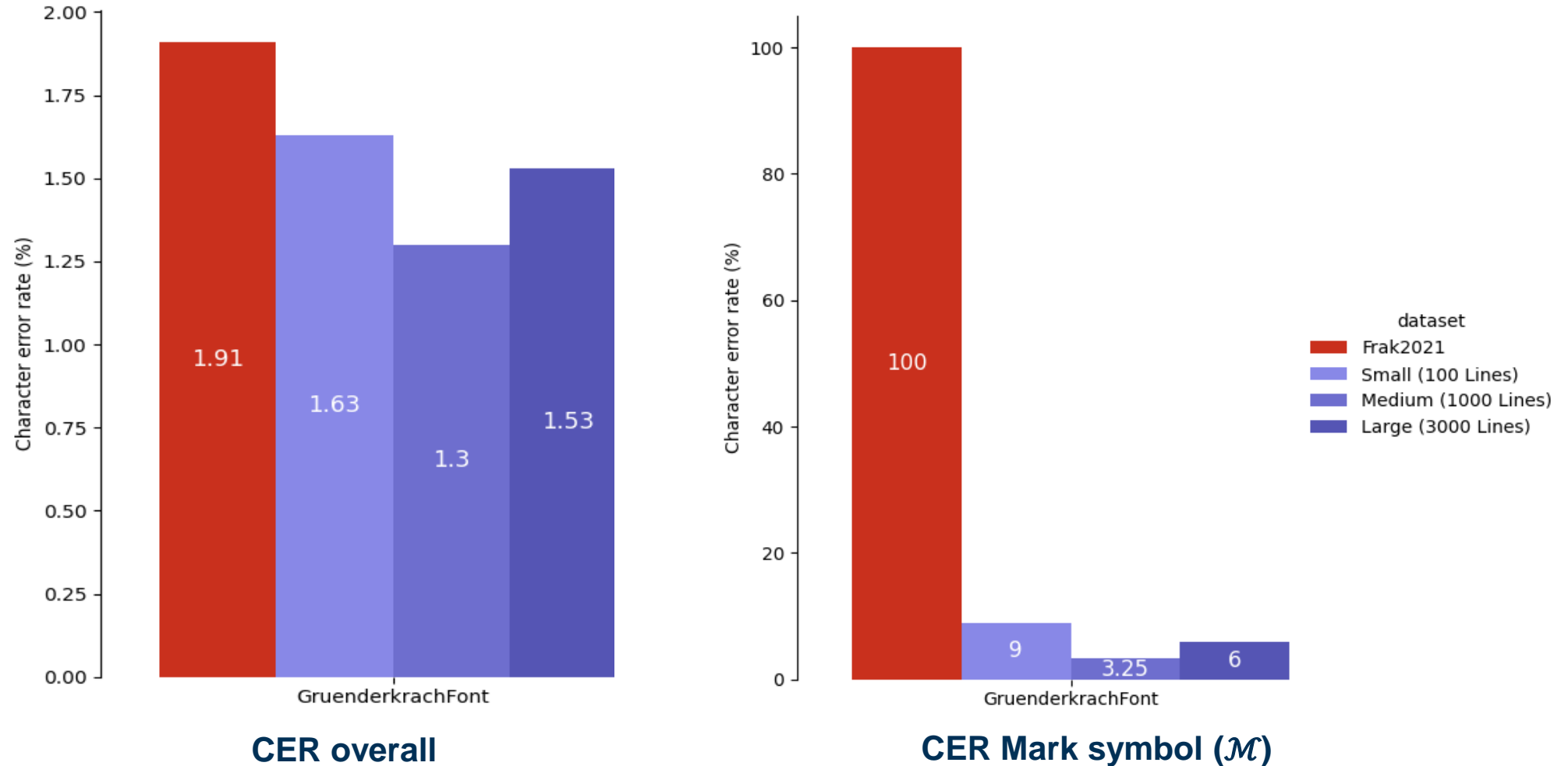


Fig. 1

### 3. Training with synthetic ground truth (Approach A)

Evaluation: **Berthold Mainzer Fraktur**

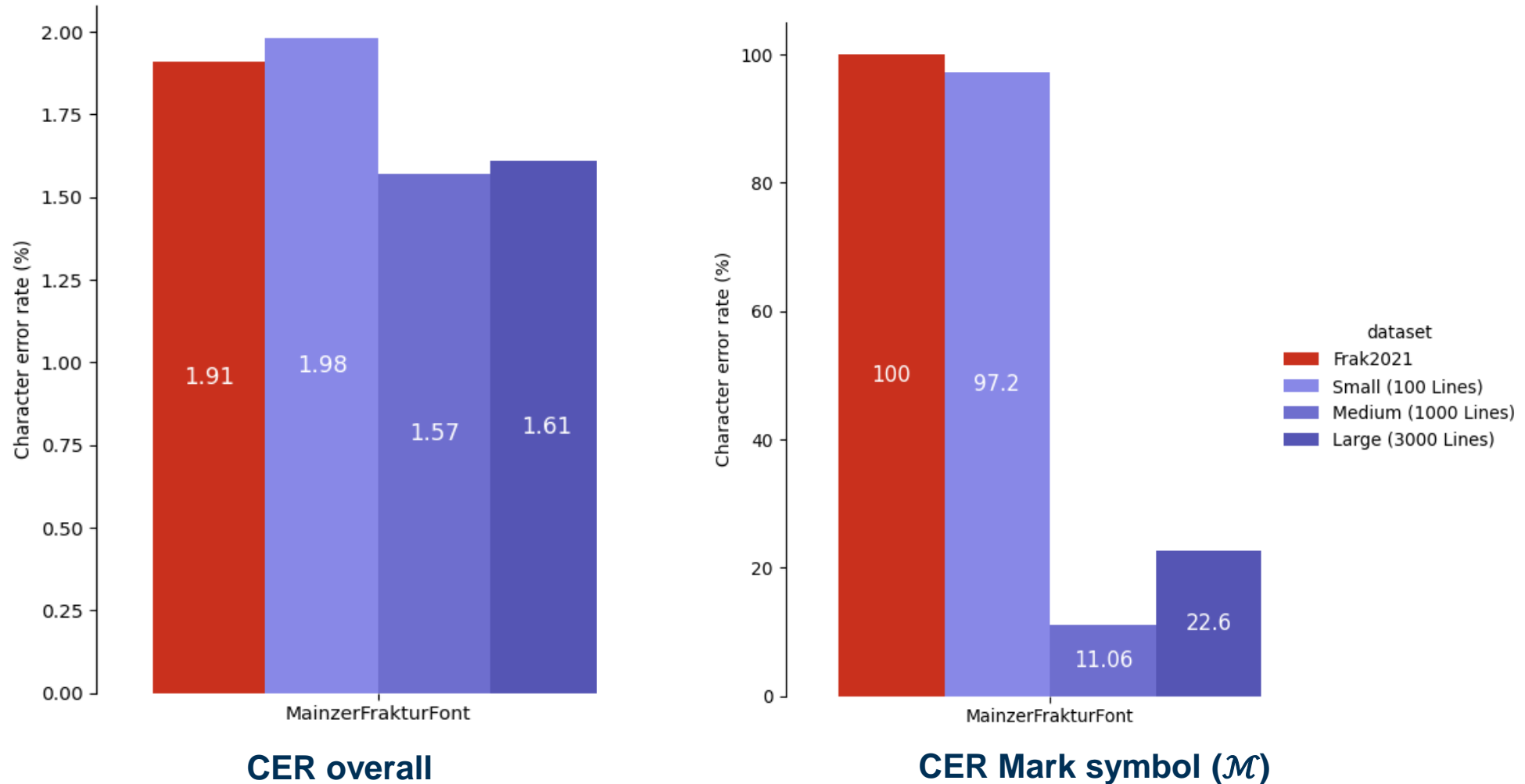


Fig. 2

# Finetune your OCR! Improving automated text recognition for early printed works by finetuning existing Tesseract models



1. OCR and work-specific training at UB Mannheim
2. Introduction
3. Training with synthetic ground truth (Approach A)
4. **Training with real-world ground truth (Approach B)**
5. Conclusion

## 4. Training with real-world ground truth (Approach B)



- like the synthetic training, the source material for our ground truth production was the "Charlottenburger Amtsschrifttum", 1879-1919 (Zentral- und Landesbibliothek Berlin)



## 4. Training with real-world ground truth (Approach B)



- like the synthetic training, the source material for our ground truth production was the "Charlottenburger Amtsschrifttum", 1879-1919 (Zentral- und Landesbibliothek Berlin)
- the transcription of the source material was done in Transkribus

## 4. Training with real-world ground truth (Approach B)

- like the synthetic training, the source material for our ground truth production was the "Charlottenburger Amtsschrifttum", 1879-1919 (Zentral- und Landesbibliothek Berlin)
- the transcription of the source material was done in Transkribus
- in contrast to the synthetic training, only a grayscale normalization for the image contents was performed as data augmentation

## 4. Training with real-world ground truth (Approach B)

- like the synthetic training, the source material for our ground truth production was the "Charlottenburger Amtsschrifttum", 1879-1919 (Zentral- und Landesbibliothek Berlin)
- the transcription of the source material was done in Transkribus
- in contrast to the synthetic training, only a grayscale normalization for the image contents was performed as data augmentation
- both number of ground truth lines as well as training runtime (in epochs) were identical to synthetic training

## 4. Training with real-world ground truth (Approach B)

Table 4: Training sets with real-world ground truth

### Work-specific training with real-world ground truth (3 Datasets)

Base model:	frak2021 (UB Mannheim)		
GT Source:	„Charlottenburger Amtsschrifttum“, 1881–1906 (ZLB)		
Ground Truth:	<b>LARGE*</b> 3.000 GT lines	<b>MEDIUM†</b> 1.000 GT lines	<b>SMALL†</b> 100 GT lines
	PNG, grayscale (+ grayscale normalization)		
Target:	Extension of the character set: $\mathcal{M}$		
Runtime:	60 epochs with termination criterion → Actual runtime approx. 40–50 epochs		

\* set contains lines with and without symbol  $\mathcal{M}$ ; † set contains only lines with symbol  $\mathcal{M}$

## 4. Training with real-world ground truth (Approach B)

Evaluation: real-world ground truth

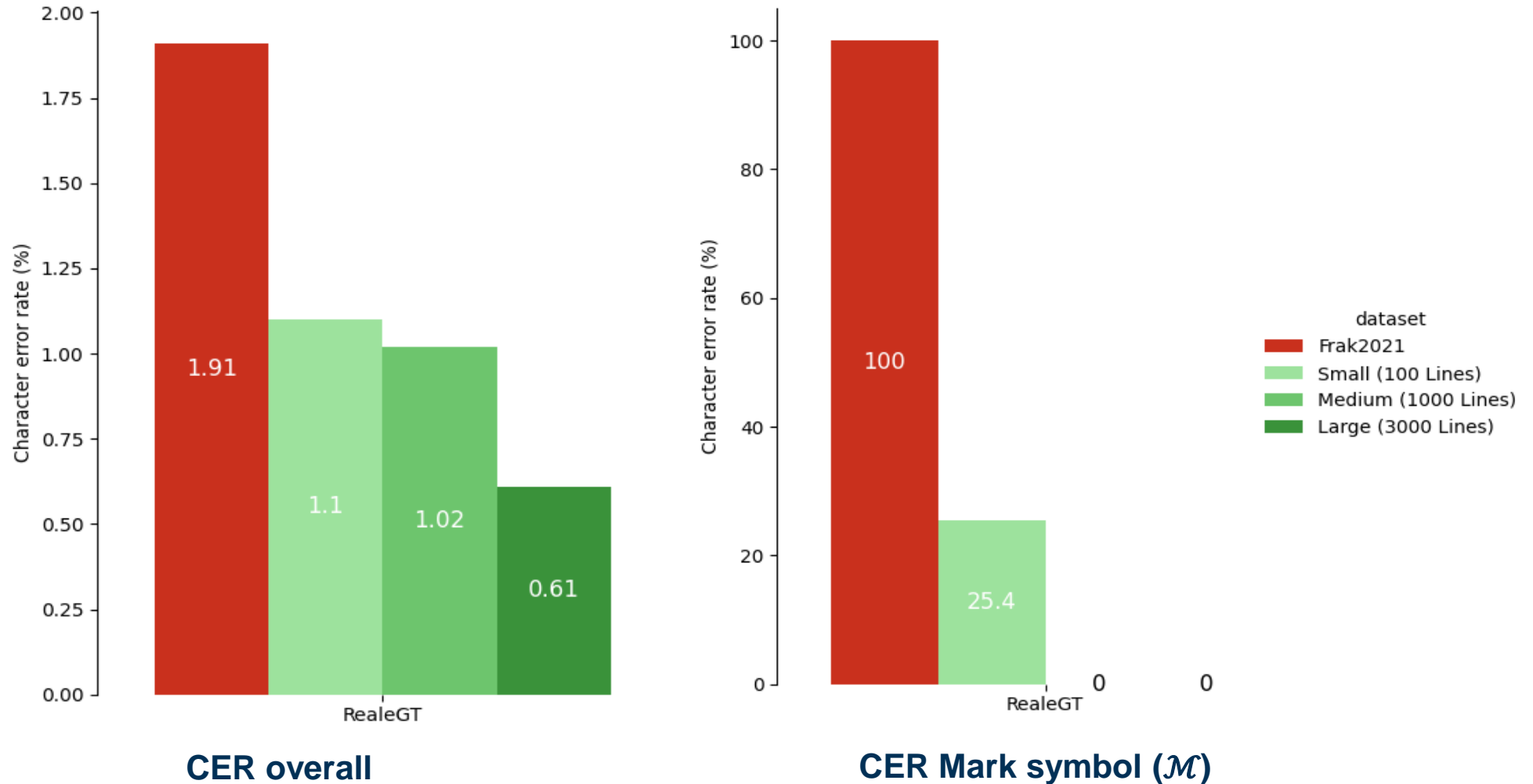


Fig. 3

# Finetune your OCR! Improving automated text recognition for early printed works by finetuning existing Tesseract models



1. OCR and work-specific training at UB Mannheim
2. Introduction
3. Training with synthetic ground truth (Approach A)
4. Training with real-world ground truth (Approach B)
5. **Conclusion**

## 5. Conclusion

### CER overall (synthetic and real-world GT)

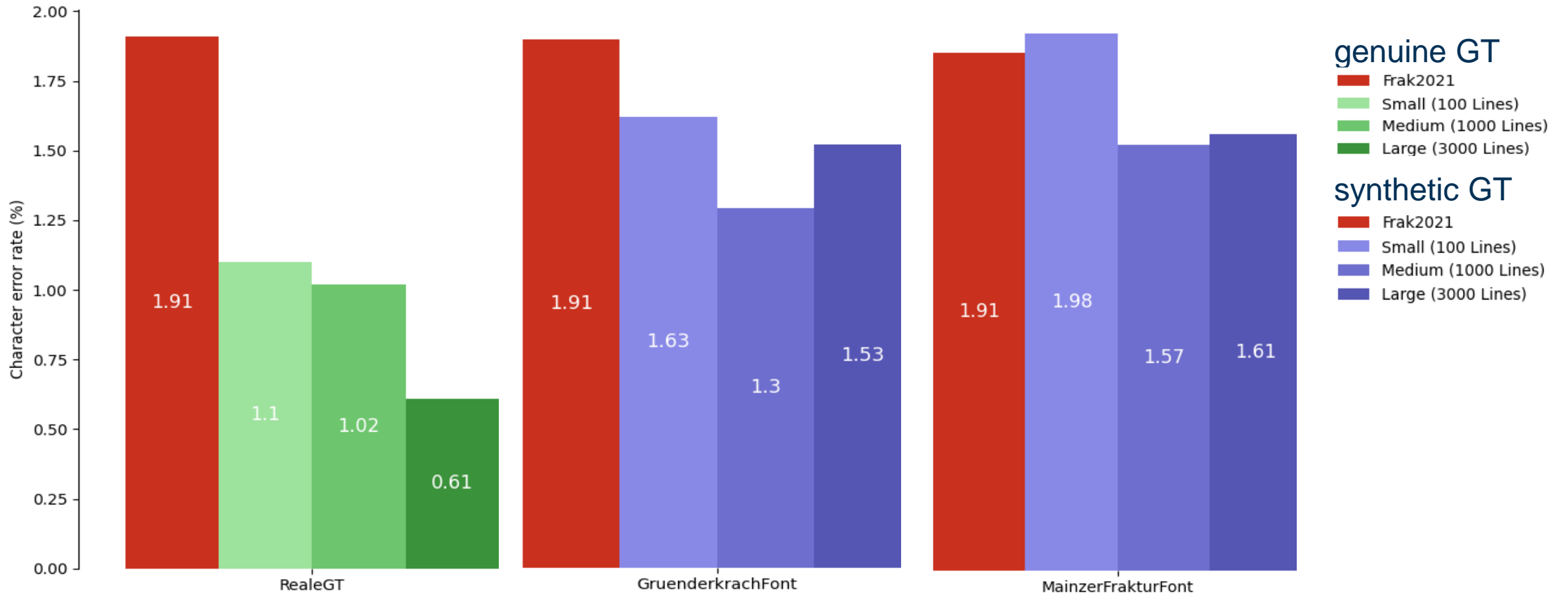


Fig. 4

## 5. Conclusion

CER Mark symbol  $\mathcal{M}$  (synthetic and real-world GT)

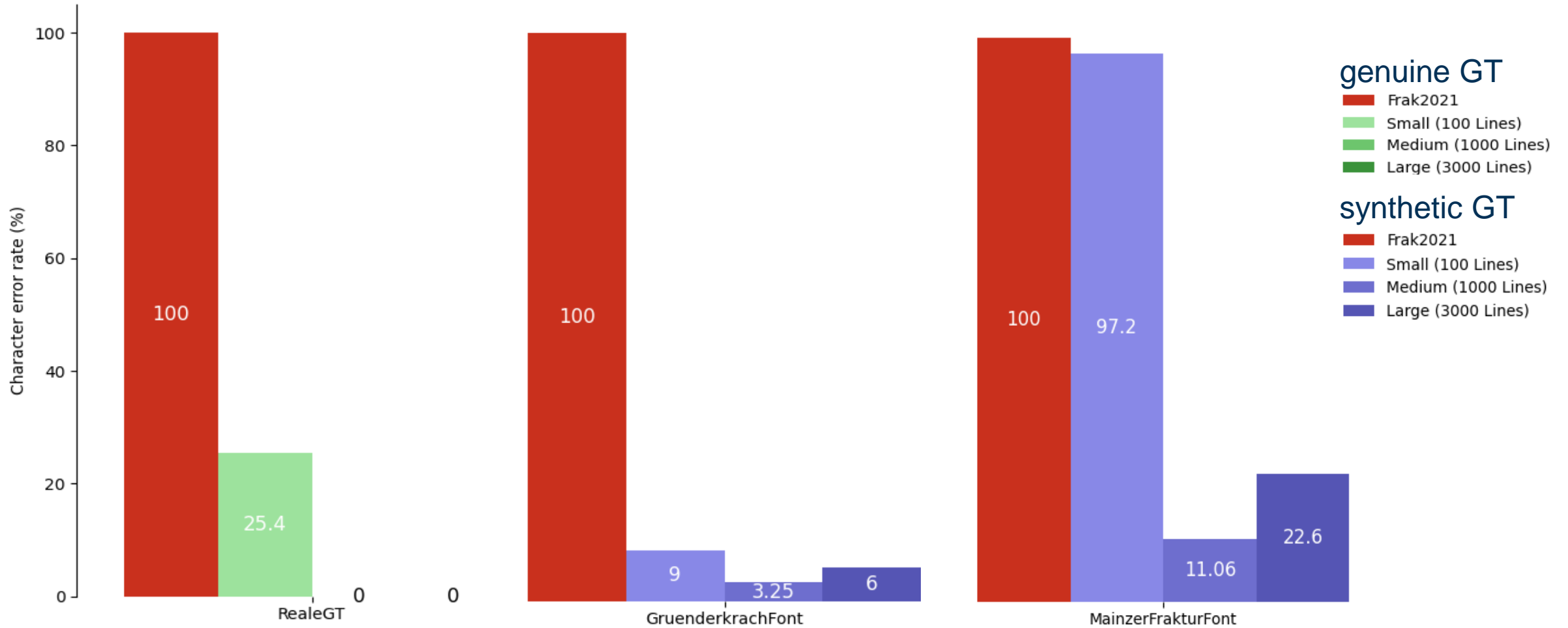


Fig. 5



## 5. Conclusion

Training with synthetic ground truth		Training with real-world ground truth	
Advantage	Disadvantage	Advantage	Disadvantage
Low time required to generate extensive GT sets	Fonts corresponding to source material are rarely available	Very good improvement of the base model	Relatively high effort to create the GT (selection of corresponding text passages with eligible glyphs + transcription + quality assurance)
Very good results when used font matches source material	Character set of available fonts mostly insufficient (especially for rare glyphs)	Trained model is superior to trained models with synthetic ground truth	
	Rare glyphs inconsistently encoded in different fonts		
	In general, fonts do not include glyph variations		

## 5. Conclusion

**Berthold Mainzer Fraktur** (synthetic ground truth):

Ende März 1892 auf 25 587 525,82 *M* gegen 23 606 652,22 *M* Ende

**Gründerkrach Fraktur** (synthetic ground truth):

Ende März 1892 auf 25 587 525,82 *M* gegen 23 606 652,22 *M* Ende

**„Charlottenburger Amtsschrifttum“** (real-world ground truth):

Ende März 1892 auf 25 587 525,82 *M*. gegen 23 606 652,22 *M*. Ende  
März 1891 ...

## 5. Conclusion

„**Charlottenburger Amtsschrifttum**“ (real-world ground truth): more  $\mathcal{M}$  variants

1250  $\mathcal{M}$  auf 937,50  $\mathcal{M}$

121 280  $\mathcal{M}$ , d. h. rund 1850  $\mathcal{M}$

25 000  $\mathcal{M}$ , also um

1125  $\mathcal{M}$ , sondern 750  $\mathcal{M}$ .

## 5. Conclusion

- Finetuning is an iterative and adaptive modern approach
- Besides pre- and post-processing, finetuning is an important tool to optimize OCR results
- Even a few hundred lines of ground truth can have a huge impact on training results
- Other projects following this approach:
  - Transkribus
  - eScriptorium
  - OCR4All
  - PERO-OCR

# Finetune Your OCR!

# Finetune your OCR! Improving automated text recognition for early printed works by finetuning existing Tesseract models



Ground truth data, further evaluations and additional information on GitHub:  
<https://github.com/UB-Mannheim/charlottenburger-amtsschrifttum>

Thank you!

**Jan Kamlah (Development):** [jan.kamlah@bib.uni-mannheim.de](mailto:jan.kamlah@bib.uni-mannheim.de)

**Thomas Schmidt (Project management):** [thomas.schmidt@bib.uni-mannheim.de](mailto:thomas.schmidt@bib.uni-mannheim.de)

## References

- „Charlottenburger Amtsschrifttum“, 1879–1919 (ZLB),  
<https://digital.zlb.de/viewer/metadata/16046633/1/>
- BertholdMainzerFraktur, Peter Wiegel,  
<http://www.peter-wiegel.de/MainzerFraktur.html>
- Weil, Stefan: Neue Frakturmodelle für Tesseract, 2019,  
<https://madoc.bib.uni-mannheim.de/53748>
- Weil Stefan: 126 Jahre Zeitung online - Fundgrube für historisch Interessierte und Motor für die Bibliotheks-IT : 126 years of the newspaper online, 2018,  
<https://madoc.bib.uni-mannheim.de/46507/>
- Weil, Stefan und Kamlah, Jan: Forschungsdaten aus Digitalisaten, 2019,  
<https://madoc.bib.uni-mannheim.de/52204>