Report: A Transkribus HTR-project for the recognition and transcription of Konrad Zuse's handwriting

Author: Johannes Dickhaut

Course: Semantic Digital Libraries – Università di Bologna

Prof. Giovanni Colavizza

Date: 06.01.2026

Chapter 1 – Introduction

This report presents the results of an experiment that attempted to train a handwritten text recognition (HTR) model to recognize and transcribe Konrad Zuse's[1] handwriting using the *Transkribus* software[2]. The author came up with the idea for the first time during the research for his bachelor's thesis, which he dedicated to the German computer pioneer. The project at that time required extensive transcription and analysis of Zuse's correspondence. Already at that time, the author considered using Transkribus, but ultimately decided against it. There were many reasons for this, primarily the uncertainty if such an experiment would produce useful results in a short period of time.

As part of the *Semantic Digital Libraries* course, the author became aware of the *Transkribus* software again and decided to venture into the project he had once rejected. The aim was to conduct a pilot experiment to test the extent to which it is possible to train a recognition model optimized for Zuse's handwriting. From the start, it was clear that it would be unrealistic to produce a fully operational model as a result. Rather, the work process and the results as a whole were to be critically evaluated in order to discuss and consider potential practical applications, for example in an archive.

To ensure a clear presentation of the results, this report is structured as follows: First, Chapter 2 describes the organizational steps that were necessary to obtain Zuse's handwritten documents. These documents formed the training material for the HTR-model training. Chapter 3 describes the annotation campaign process. Chapter 4 goes into more detail about the results of the annotation campaign and the composition of the training material and evaluates them critically. Chapter 5 then focuses on the model training process and its results. Finally, Chapter 6 discusses whether training a Zuse handwriting model could contribute to accelerating the digitization of existing material in practice.

Chapter 2 – The collection of Zuse's handwritten material

The first phase of the project involved collecting as much training material as possible. In this case, the potential training material consisted of documents handwritten by Konrad Zuse, which then had to be transcribed. The point of contact for this was the archive of the Deutsches Museum (DMA) in Munich[3], which manages Zuse's estate. The estate comprises a total of over 4.000 items/signatures, whereby one item can consist of several documents. At this point, the question arose as to how Zuse's handwritten documents could be located within this large amount of material. For purposes like this, the DMA

[1] https://en.wikipedia.org/wiki/Konrad_Zuse
[2] https://www.transkribus.org/
[3] https://digital.deutsches-museum.de/de/digital-catalogue/archive-unit/NL%20207/

provides an XML data set online in which the contents of the estate are listed in a structured manner according to *Encoded Archival Description (EAD)* guidelines[4]. Using an XQuery (Figure 1), the author was able to get the signatures of handwritten documents.

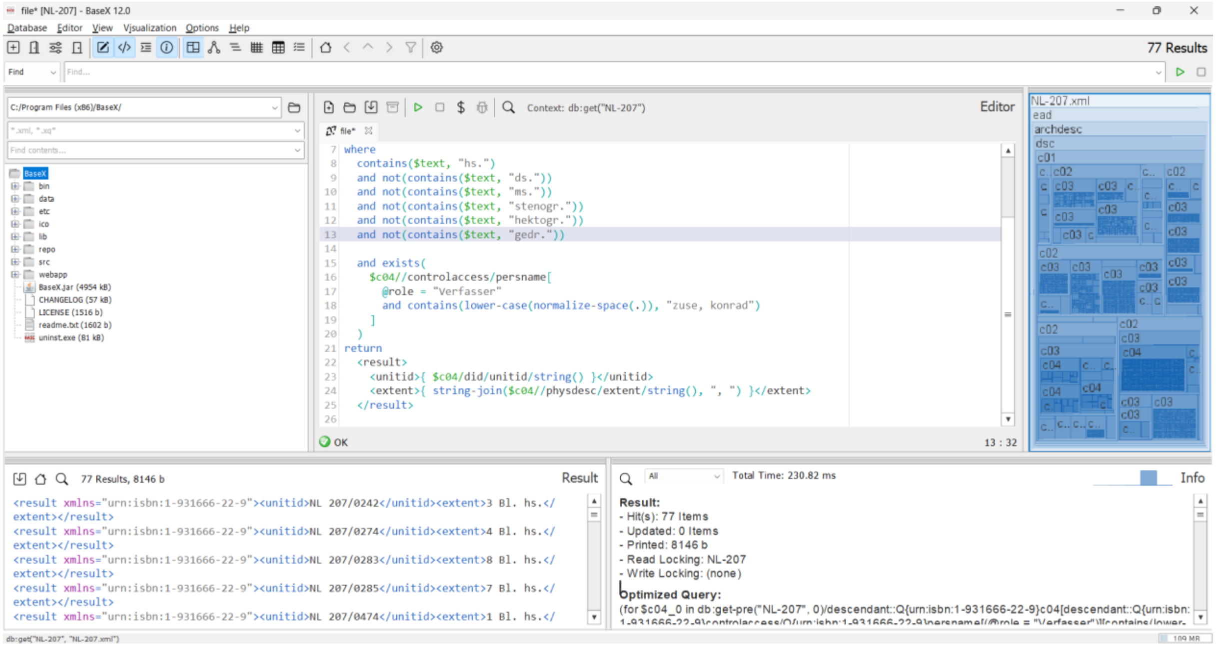


Figure 1: XQuery for the filtering of Zuse's hanwritten documents..

It should be noted that selecting only on the basis of the attribute “hs.” (handschriftlich = handwritten) would not have been sufficient. Zuse wrote most of his notes in shorthand style. Therefore, the XQuery had to be adjusted to exclude these search results. The query returned overall 77 results. Some of them have been made available by the DMA via Deutsches Museum Digital. However, the resolution quality of these online documents was not sufficient for Transkribus model training. The author therefore had to contact the archive management to obtain the desired holdings in suitable quality. This worked without any further problems.

However, during the examination of the material, it became clear that a large portion of it was flawed. Some documents still contained shorthand pages. Others were drawings with only minor written additions. Some holdings had not even been written by Zuse himself. All of these erroneous results were not the product of a faulty XQuery, but were due to inaccuracies and errors in the inventory overview in the XML data set. Furthermore a renewed search of the online holdings identified further signatures that contained additional handwritten documents. These were requested from the archive management in a second attempt and were made available.

[4] https://www.loc.gov/ead/

Chapter 3 – Annotation Campaign

Before the actual annotation campaign began, the material was carefully reviewed once again and unsuitable documents were removed. These included, for example, letters written by Zuse up until the 1930s. His handwriting during this period differed greatly from his later handwriting (see Figures 2 and 3). Pages with smudged ink and extensive blackouts/crossings-out were also omitted.

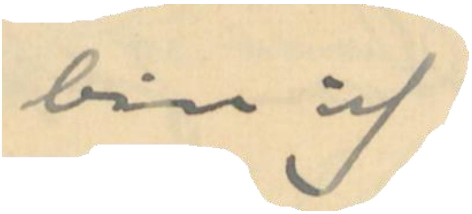


Figure 2: "bin ich" Zuse handwriting 1930 (DMA, NL 207, 1118, p. 1).

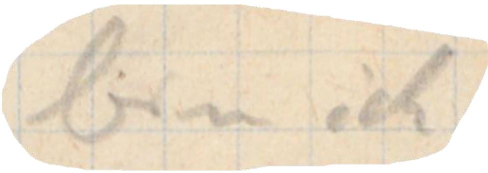


Figure 3: "bin ich" Zuse handwriting 1945 (DMA, NL 207, 1131, p. 1).

The actual transcription and annotation of the documents was based on the *OCR-D Annotation Guidelines*[5]. This meant that the text was transcribed verbatim, even if Zuse had included spelling mistakes. The tagging function of Transkribus was used in various ways. Firstly, it was used to mark underlining, strikethroughs, and superscript/subscript characters (italic or bold elements were not tagged, as the handwritten differences were not clear enough). Passages that were unclear or blurred were marked with the unclear tag so that they would not be included in the training. Furthermore, the people mentioned in the documents were linked to their Wikidata entries using the person tag to enable clear identification. Due to time constraints places, organizations, etc. were not additionally tagged. If a scientific edition is to be created in the future based on the transcribed material, this would have to be done in order to ensure the clarity of the content and to enable semantic queries.

All in all, it became clear during this project phase that early hands on experience with the subject matter is essential. One obvious hurdle is Zuse's difficult handwriting style. If the author had not already worked with the texts in the course of his bachelor's thesis and become used to the handwriting, this step would probably have taken considerably more time. In addition, prior knowledge of the subject matter was an advantage, as it made it easier to identify some of the technical terms used by Zuse in his texts.

[5] https://ocr-d.de/de/gt-guidelines/trans/transkription.html

Chapter 4 – The training material as a result of the annotation campaign

All in all, 72 pages were transcribed, each with varying amounts of content. The exact number of words transcribed was 8871. This meant that the amount of transcribed material was just below the 50 pages or 10,000 words recommended by Transkribus for handwriting training. Due to the problems described in Chapter 2 regarding the collection of material, time pressure, and the pilot project nature of the experiment, which was clear from the start, it was decided in this situation to remain below the recommended training material limit and to not transcribe any further, potentially unsuitable documents.

Regardless of this, the representativeness of the training sample had to be ensured. The documents therefore covered different font sizes and pen types (fountain pen, pencil, ballpoint pen). Transkribus itself did not offer suitable functionalities for more detailed analyses of representativeness. To solve this problem and obtain more accurate information on the transcription content, the author created a Python script[6] to check the character quantity in the corpus. The results of the character frequencies for letters, numbers, and special characters can be seen in Figures 4, 5 and 6 (also as tables in Appendix 1, 2, 3 for a better overview).

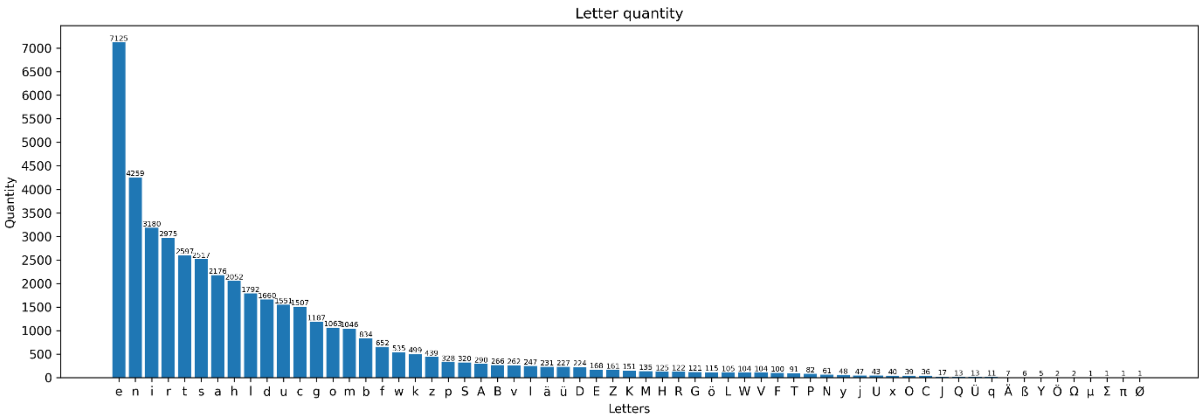


Figure 4: Letter quantity.

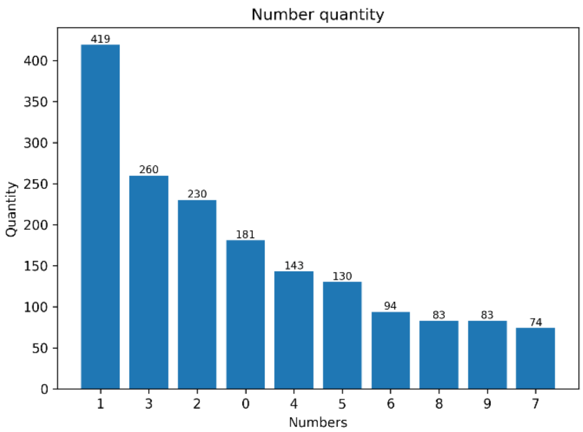


Figure 5: Number quantity.

[6] https://github.com/jojom4n98/zuse_handwriting_semantic_digital_libraries.git

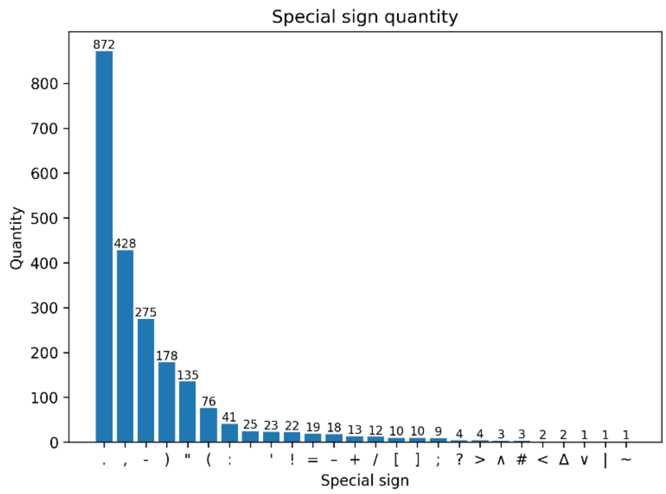


Figure 6: Special sign quantity.

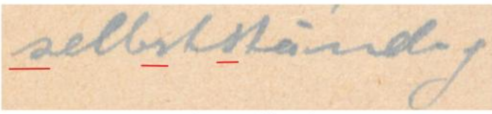
In addition, the letter frequency in Zuse's documents was compared with the general letter frequency in German texts (Figure 7). The results clearly showed that Zuse did not use any letters comparatively more or less frequently in his texts. All percentage deviations from the standard German usage were less than 1%. Thus, in terms of character distribution, it was a representative sample.

Letter	Zuse letter usage	German letter usage	Percentage difference
e	16,53	17,4	0,87
n	9,79	9,78	-0,01
i	7,77	7,55	-0,22
r	7,02	7	-0,02
t	6,09	6,15	0,06
s	6,43	7,27	0,84
a	5,59	6,51	0,92
h	4,93	4,76	-0,17
l	4,3	3,44	-0,86
d	4,27	5,08	0,81
u	3,61	4,35	0,74
c	3,5	3,06	-0,44
g	2,96	3,01	0,05
o	2,5	2,51	0,01
m	2,68	2,53	-0,15
b	2,49	1,89	-0,6
f	1,7	1,66	-0,04

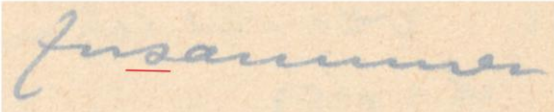
Letter	Zuse letter usage	German letter usage	Percentage difference
w	1,45	1,89	0,44
k	1,47	1,21	-0,26
z	1,36	1,13	-0,23
p	0,93	0,79	-0,14
v	0,83	0,67	-0,16
ä	0,54	-	-
ü	0,54	-	-
ö	0,27	-	-
y	0,12	0,04	-0,08
j	0,15	0,27	0,12
x	0,09	0,03	-0,06
q	0,05	0,02	-0,03
ß	0,01	0,31	0,3
ω	0	-	-
μ	0	-	-
σ	0	-	-
π	0	-	-
ø	0	-	-

Figure 7: Comparison letter usage percentage in Zuse texts and general german.

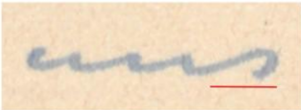
Nevertheless, it was clear that the training material was far from ideal. Some letters, such as *q*, *y* or *x* were transcribed very rarely. On this basis, successful training was unrealistic. A minimum limit for transcriptions per character cannot be specified at this point. In addition, problems arose from the cursive writing technique, as some letters are written differently in combination. Figure 8 illustrates this using the letter *s* as an example, which takes on a different form depending on whether it appears at the beginning, middle, or end of a word. Suitable training material would have to adequately cover all these eventualities. The Python script used previously was not designed for such a check.



‚s‘ at multiple positions in a word



‚s‘ in the middle of a word



‚s‘ at the end of a word

Figure 8: Different variations of the letter 's' in Zuse's handwriting.

Consequently, the amount of training material would have to be significantly higher than would have been possible in the course of this transcription project. Further considerations regarding realistic implementation as a project, e.g., for an archive, are presented in Chapter 6.

Chapter 5 – Model training and results

The model training was carried out in four rounds with different setups in order to evaluate the quality and the differences of the results. In all cases, Transkribus' standard recommendations were followed. Accordingly, 10% of the transcribed material was used as a validation set. The training process comprised a total of 100 cycles, with the possibility of termination after cycle number 20 if no further progress was measured in the Character Error Rate (CER). Lines marked with the unclear tag were omitted.

Initially, two models were trained, which were based on *existing line polygons* (possible setting in Transkribus training) during the training process. The author's idea was that this approach could help to better recognize inconsistent baselines. The model *Zuse handwritten – from scratch* (Model I)[7] was not

[7] https://app.transkribus.org/models/text/459125

trained on the basis of an existing model. The training process was run through to the 91st cycle, ultimately achieving a Training CER (TCER) of approximately 20.5% and a Validation CER (VCER) of 36.24% (Figure 9).

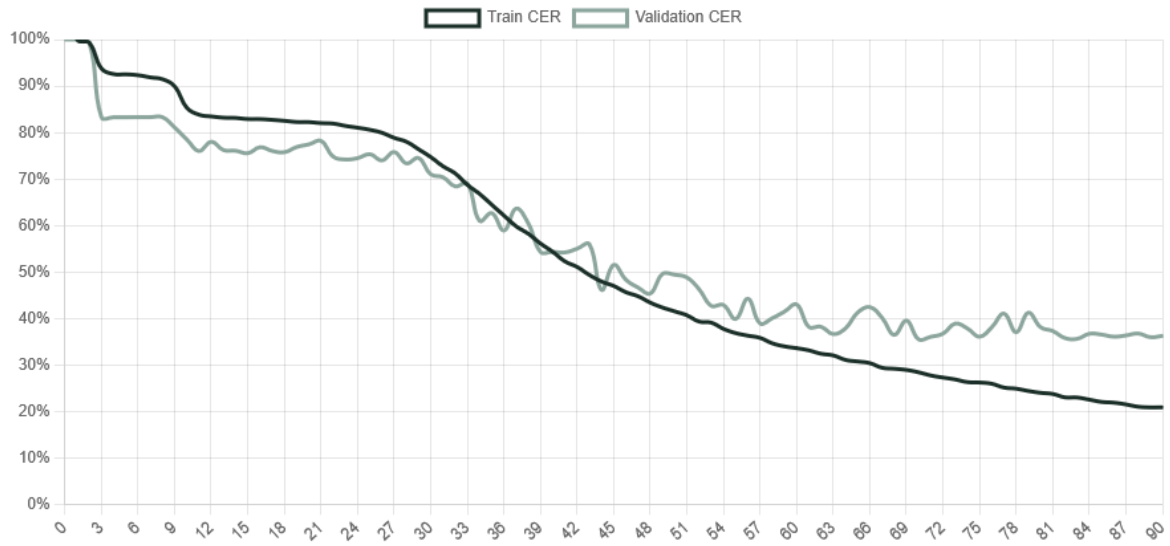


Figure 9: TCER and VCER curves 'Zuse handwritten – From scratch' (model I).

The *Zuse handwritten – tuned base model* (Model II)[8] was also not trained on the basis of existing line polygons. In this case, the Transkribus model *German Giant I* was used as the base HTR model. After completing all 100 training cycles, it had a TCER of approximately 24.82% and a VCER of 31.72% (Figure 10).

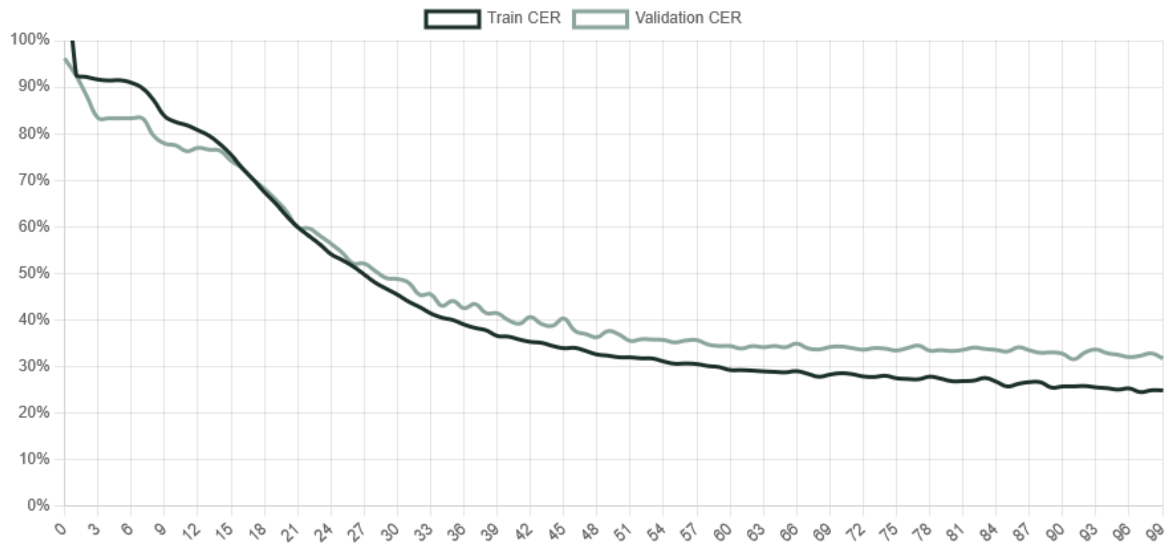


Figure 10: TCER and VCER curves 'Zuse handwritten – tuned base model' (model II).

Both Model I and Model II showed a clear overfitting effect. Compared to existing models such as the supermodel *Text Titan I ter* or *German Giant I*, both models performed poorly and would not have

[8] https://app.transkribus.org/models/text/459305

delivered satisfactory results. The fact that Model II performed worse than *German Giant I* did not meet previous expectations, as the training was aimed at improving the existing model.

In the second training phase of the project, the approach was to train models not based on existing line polygons. Like Model I, the model *Zuse handwritten – From scratch without line polygons* (Model III)[9] was not trained based on an existing model. The training spanned all 100 cycles. In the end, the model had a VCER of 16.68% and a TCER of approximately 17.06%. Also Model III developed an overfitting effect. As in Model I, the course of the VCER curve contained large deviations from cycle to cycle (Figure 11).

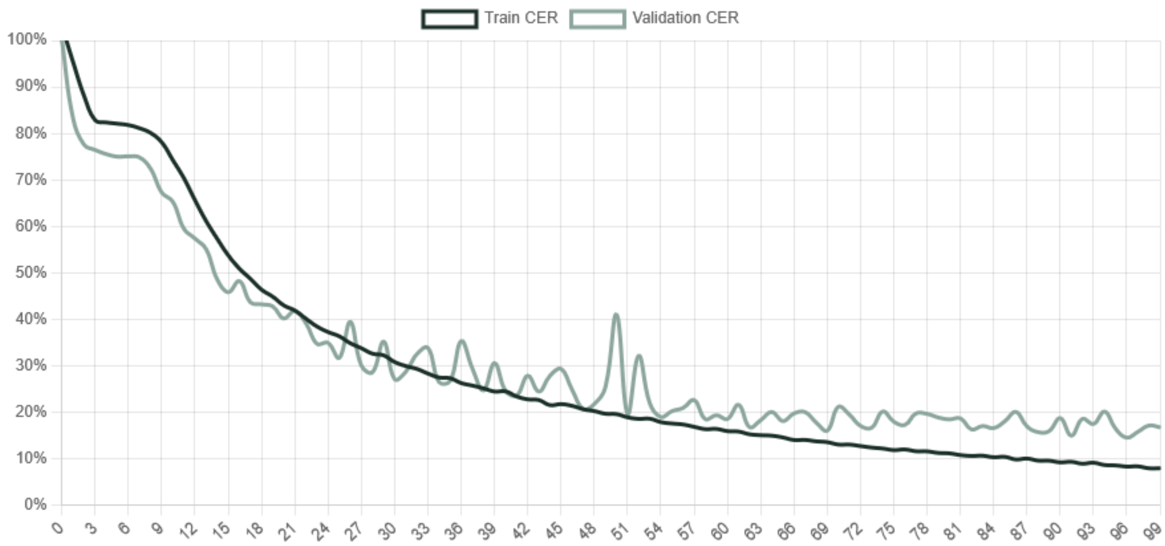


Figure 11: TCER and VCER curves 'Zuse handwritten – From scratch without line polygons" (model III).

The model *Zuse handwritten – tuned base model without line polygons* (Model IV)[10] achieved the best results in the project phase. After 100 training cycles, it had a VCER of 10.73% and a TCER of approximately 10.73%. In contrast to the other models, no overfitting effect was apparent here according to the curve. The fluctuations within the VCER curve were minor (Figure 12).

[9] https://app.transkribus.org/models/text/464445
[10] https://app.transkribus.org/models/text/463085

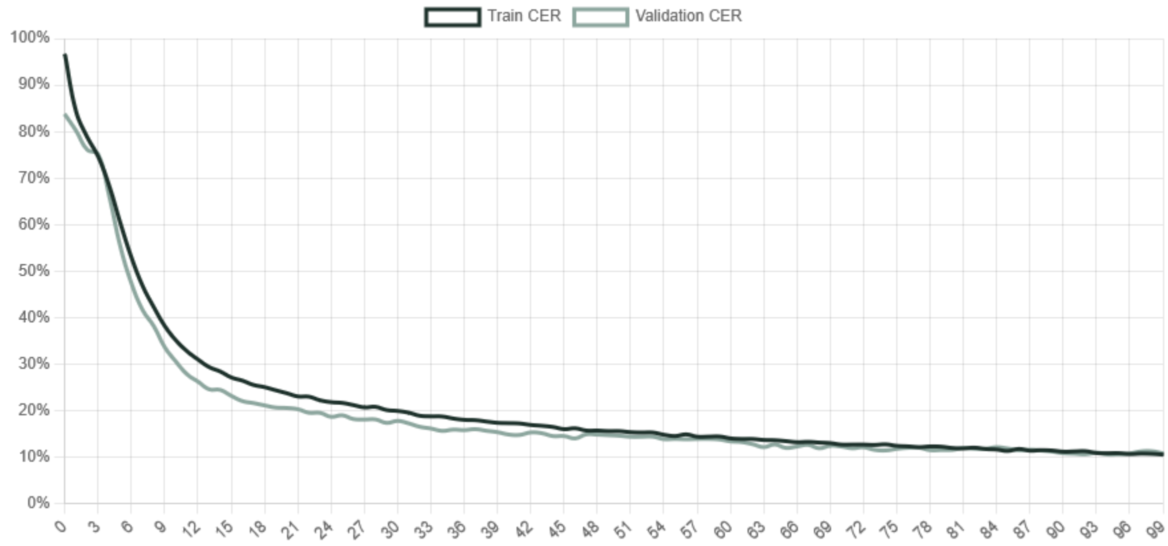


Figure 12: TCER and VCER curves ‚Zuse handwritten – tuned base model without line polygons' (model IV).

Figure 13 shows a comparison of the transcription results of Model IV, *German Giant I* and *Text Titan I ter* using an example document that was part of the validation set from Model IV training. In this case, Model IV produced a result that surpassed that of *German Giant I* and was even on par with that of *Text Titan I ter*.

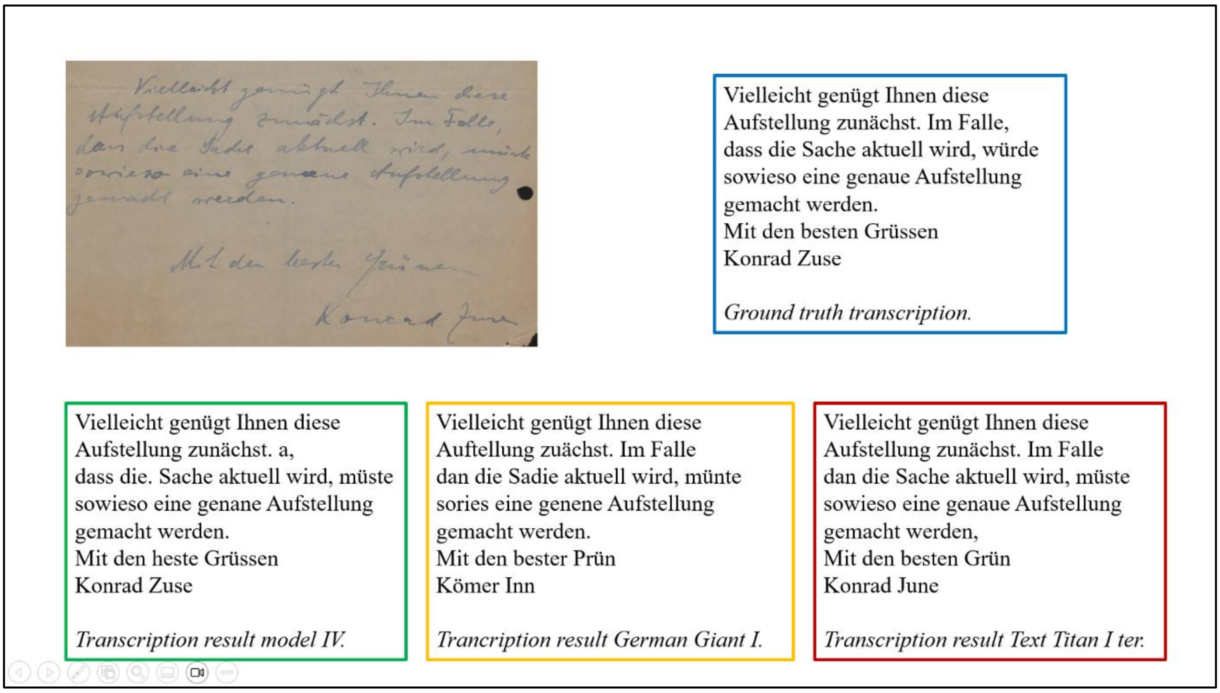


Figure 13: Comparison transcription results of a document from model IV validation set.

After training was complete, Model IV was manually tested with additional documents that were not part of the validation/training set in order to verify its potential practical application.

It became clear that the model would only be of limited use in practice, as it has significant problems with document layout and line recognition in addition to imperfect character recognition. Figure 14 shows an example where such problems occurred. The red numbers correspond to the recognized document regions. It is clear that the recognition of individual passages continues to be at a comparatively high level. However, reconfiguring the individual layout regions would take a lot of time in practice. The models *German Giant I* and *Text Titan I ter* have similar problems.

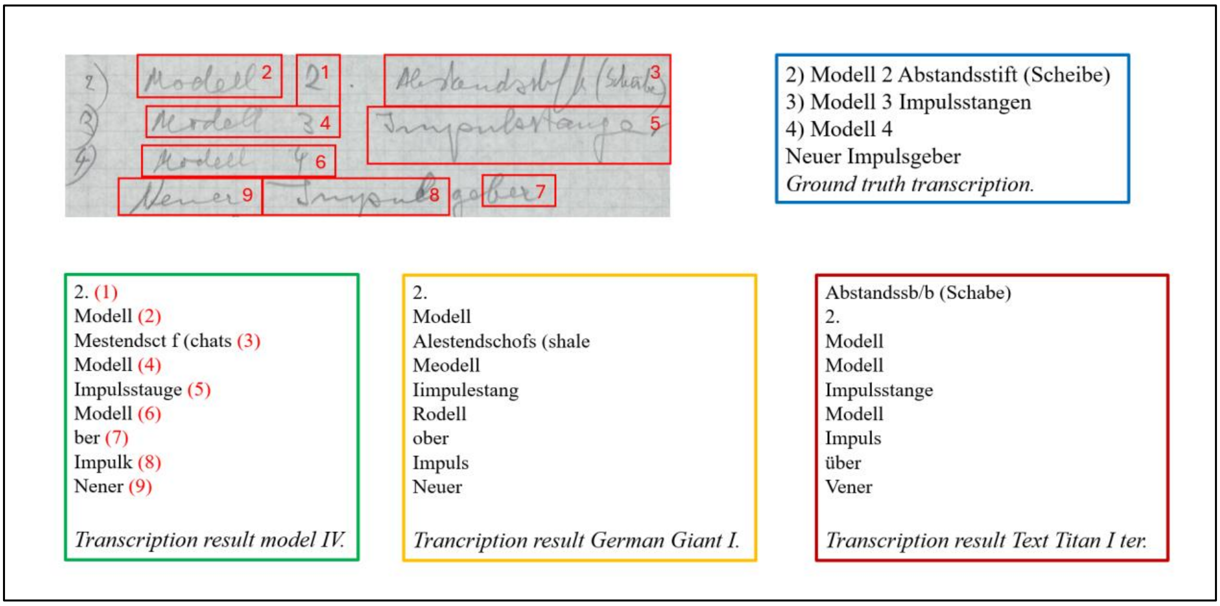


Figure 14: Comparison transcription results of a document with problems regarding baseline detection.

Figure 15 shows another example where the layout and baseline recognition for all three models was so poor that only a fraction of the text was processed. This would probably occur more frequently with lower-quality documents.

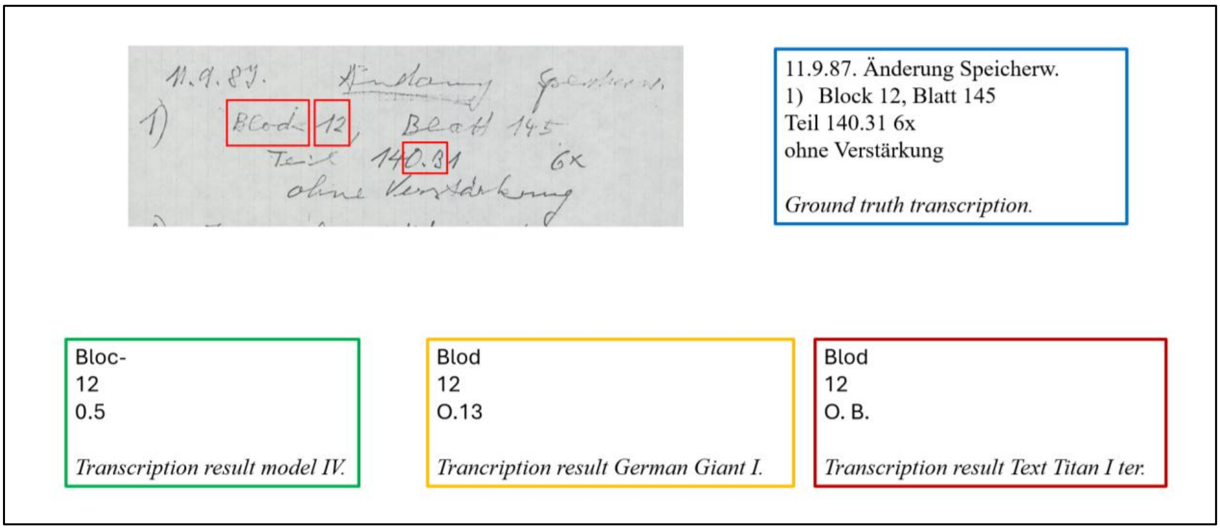


Figure 15: Comparison transcription results of a document with significant problems regarding baseline detection.

Chapter 6 – Potential use in practice

As became clear in Chapter 5, the results were an improvement compared to already existing Transkribus HTR models but still far from optimal, meaning that even minor improvements within the scope of the training parameters/annotation campaign would most likely not have produced significantly better results. Therefore, the realistic implementation of such a project in practice should be questioned.

In the author's opinion, the digitization of the handwritten material in Zuse's estate could be partly but not significantly accelerated by training a specific handwriting model. Although Zuse's estate is extensive, the proportion of handwritten documents would not provide sufficient training material to achieve satisfactory results. By the time this level of accuracy was achieved, almost all handwritten documents would probably have been transcribed. From a project management perspective, it should also be noted that the organizational overhead would be significantly higher than in the present project. New employees would first have to get familiar themselves with Zuse's handwriting and learn about the subject matter. Later, extensive comparisons of the individual transcriptions would be necessary.

Model IV could be theoretically used for individual passages during the digitization and achieve an acceleration effect. In other passages, however, the susceptibility to errors in terms of layout and baselines would be so high that it would be impossible to predict how extensive the adjustments to the outputs would have to be in order to produce results that could be further processed.

Chapter 7 – Conclusion

This report is the result of an experiment that attempted to optimize a Transkribus model for transcribing Konrad Zuse's handwriting. During the experiment process, it became clear what challenges such a project poses. Despite extensive support from the DMA in obtaining materials, the first problems arose early on, as some of the documents were not suitable as training material. The actual transcription process was also very time-consuming, despite the author's prior knowledge. The results produced were an improvement (Model IV) but for the usage in practice still unsatisfactory. This led to the conclusion that such an approach would not be effective in the context of digitizing Zuse's estate. It should be noted at this point that this assessment cannot be applied to other practical examples. There are certainly estates of other historical personalities that consist of significantly more handwritten material and whose handwriting is easier to decipher. However, even in these cases, it should be evaluated at an early stage whether handwriting recognition training is realistic and effective.

Appendix I – Letter count

Letter	Count
e	7125
n	4259
i	3180
r	2975
t	2597
s	2517
a	2176
h	2052
l	1792
d	1660
u	1551
c	1507
g	1187
o	1063
m	1046
b	834
f	652
w	535
k	499
z	439
p	328
S	320
A	290
B	266
v	262
I	247
ä	231
ü	227
D	224
E	168
Z	161

Letter	Count
K	151
M	135
H	125
R	122
G	121
ö	115
L	105
W	104
V	104
F	100
T	91
P	82
N	61
y	48
j	47
U	43
x	40
O	39
C	36
J	17
Q	13
Ü	13
q	11
Ä	7
ß	6
Y	5
Ö	2
Ω	2
µ	1
Σ	1
π	1
Ø	1

Appendix II – Number count

Number	Count
1	419
3	260
2	230
0	181
4	143
5	130
6	94
8	83
9	83
7	74

Appendix III – Special sign count

<table>
<tr><th>Sign</th><th>Count</th><th>Sign</th><th>Count</th></tr>
<tr><td>.</td><td>872</td><td>/</td><td>12</td></tr>
<tr><td>,</td><td>428</td><td>[</td><td>10</td></tr>
<tr><td>-</td><td>275</td><td>]</td><td>10</td></tr>
<tr><td>)</td><td>178</td><td>;</td><td>9</td></tr>
<tr><td>"</td><td>135</td><td>?</td><td>4</td></tr>
<tr><td>(</td><td>76</td><td>></td><td>4</td></tr>
<tr><td>:</td><td>41</td><td>Λ</td><td>3</td></tr>
<tr><td></td><td>25</td><td>#</td><td>3</td></tr>
<tr><td>'</td><td>23</td><td><</td><td>2</td></tr>
<tr><td>!</td><td>22</td><td>Δ</td><td>2</td></tr>
<tr><td>0</td><td>19</td><td>V</td><td>1</td></tr>
<tr><td>–</td><td>18</td><td>|</td><td>1</td></tr>
<tr><td>+</td><td>13</td><td>~</td><td>1</td></tr>
</table>