

Open Source Handwritten Text Recognition on Medieval Manuscripts using Mixed Models and Document-Specific Finetuning

**Christian Reul¹, Stefan Tomasek¹,
Florian Langhanki¹, Uwe Springmann²**

¹University of Würzburg

²LMU Munich

25.05.2022

Introduction – Model Types and Where to Use Them

- Different approaches and models for different use cases
 - Model types
 - Mixed models (MM): trained on a variety of books / manuscripts / printing types / hands / ...
 - Document-specific models: trained specifically for the recognition of a single book / ...
 - More specialized models yield better results but their creation requires more effort
 - Example use-cases
 - Mass text recognition, e.g. in libraries or archives
 - Manual correction not feasible
 - Apply MM and hope for the best
 - Preparation of a critical edition
 - 100% error-free text required → high amount of manual correction inevitable
 - Recognition → correction → training → recognition → ... until everything is transcribed
- Gray area in-between

Use-Case – Overview

Digital edition project „Die Kindheit Jesu“ (Childhood of Jesus Christ) at the University of Würzburg

- Material: Medieval manuscripts written ...
 - between 1250 and 1350 in Middle High German
 - using two very common German medieval writing styles: Gothic and Bastarda
- Goal: Comparing text variants using a configurable filter system based on different distance measures
- Some distance measures, e.g. Word Mover's Distances, can and have to be trained
- Training data needed: Non-normalized text in ...
 - (very) good but not perfect quality
 - large quantities → efficiency is key → combining mixed models and document-specific finetuning

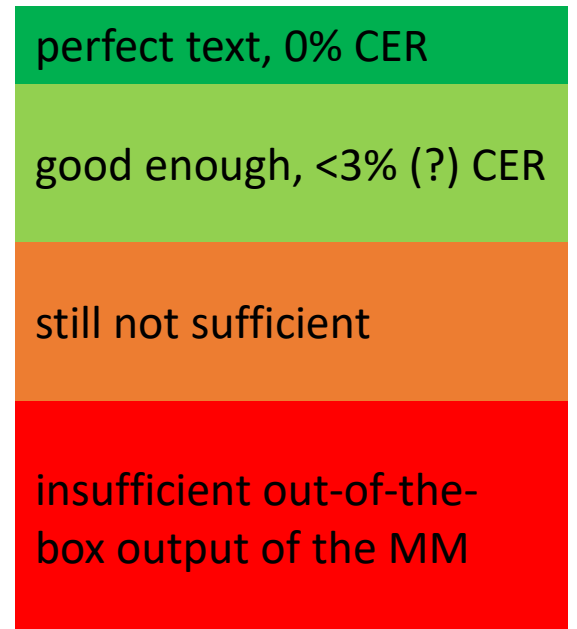
Use-Case – Task and Workflow

Task: Get from the **out-of-the-box output of the mixed model** to a **good enough CER** in a very efficient manner

Workflow:

- Train suitable mixed model(s)
- For each manuscript:
 1. Apply the most suitable MM
 2. Manually correct some pages
 3. Train a document-specific model using the newly gained GT
 4. Apply the new model
 5. Repeat steps 2-4, if needed (Iterative Training Approach)
 6. Apply the document-specific model to the remaining pages

text quality



Data – Mixed Model Training

- Suitable manuscripts from ca. 1200 to 1500: Kindheit Jesu, Marienleben, Parzival, ...
- Self-created or compiled from freely available sources
- Final corpus: 35 works (at least 30 different hands), close to 300 pages, ca. 12.5k lines



dem hovbz vñ ovgen.



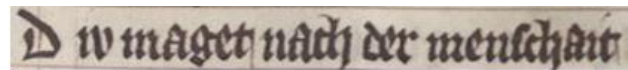
¶ Serpilleum zertribē



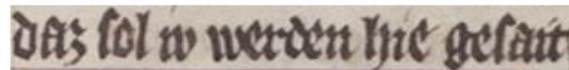
Auicena spricht wer



nüchter baden will



D iv maget nach der menschait



daz sol iv werden hie gefait



git got nit sin ewiges hi



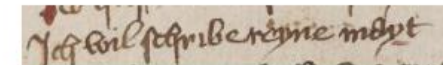
melrich Es spricht hug



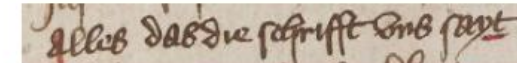
got amen Doīca t̄cia



waz ich dim. So horen



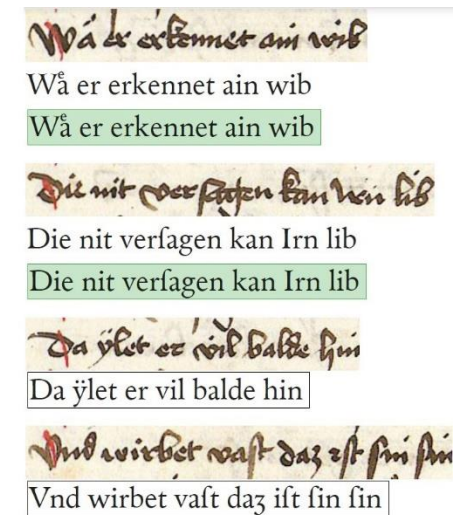
Ich wil schriben reyne mayt



Alles daz die scharfft vns sayt

Data – Preparation

- Fully open source workflow!
- Preprocessing via OCR-D (<https://github.com/OCR-D>) processors
- OCR4all and LAREX
 - Segmentation and ground truth production
 - <https://github.com/OCR4all>
- Calamari
 - Recognition, training, evaluation
 - <https://github.com/Calamari-OCR>



Calamari OCR

Data – Evaluation

- Four manuscripts (2x Gothic, 2x Bastarda) written by previously unseen hands
- Kindheit Jesu, Marienleben, Der Welsche Gast (x2)
- Each manuscript further divided into train (32 pages) and eval (8-18 pages) subsets

Daz ih des niht vol bringen chan.

Daz ih des niht vol bringen chan.

mir chom zehelfe dar an.

mir chom zehelfe dar an.

man tûn fol zu allen zÿten. vñ

man tûn fol zu allen zÿten. vñ

wâ von man nit treg fol fin vnd

wâ von man nit treg fol fin vnd

Das buch bræhte her zewege. daz sie iz alle

daz buch bræhte her zewege. daz sie iz alle

mufen lesen. die gotes kint wellent we

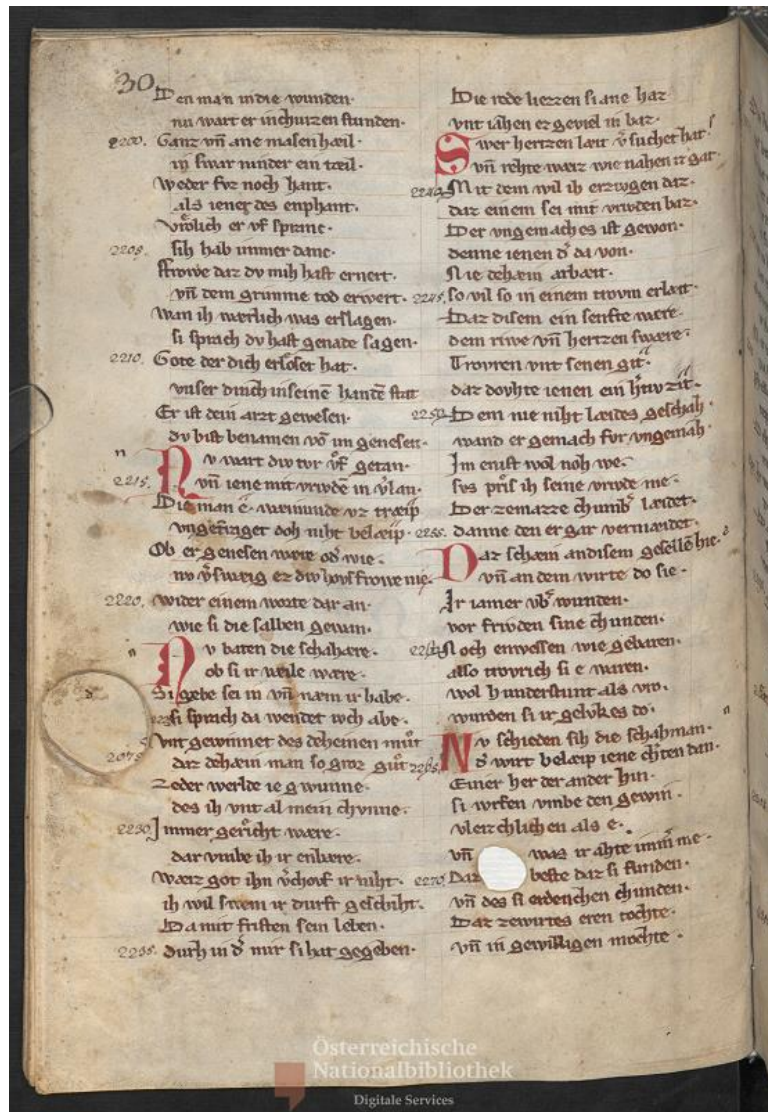
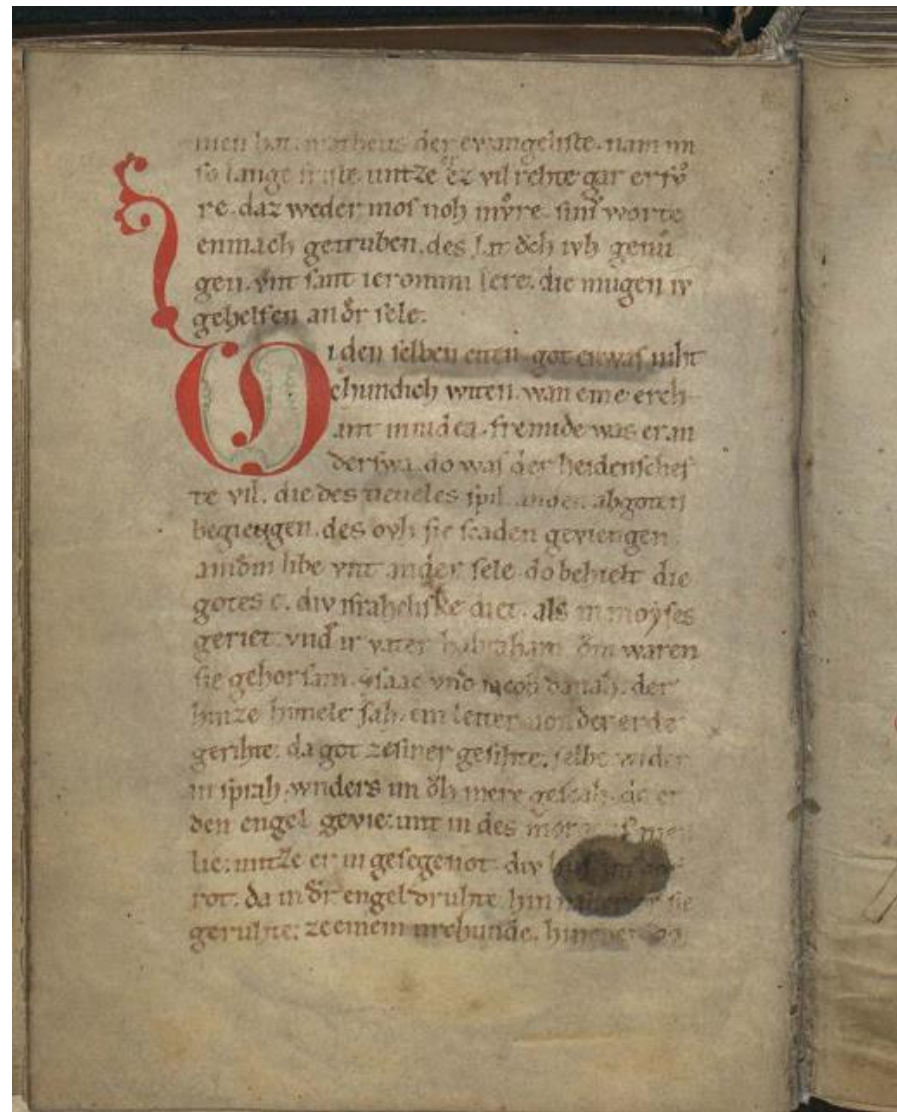
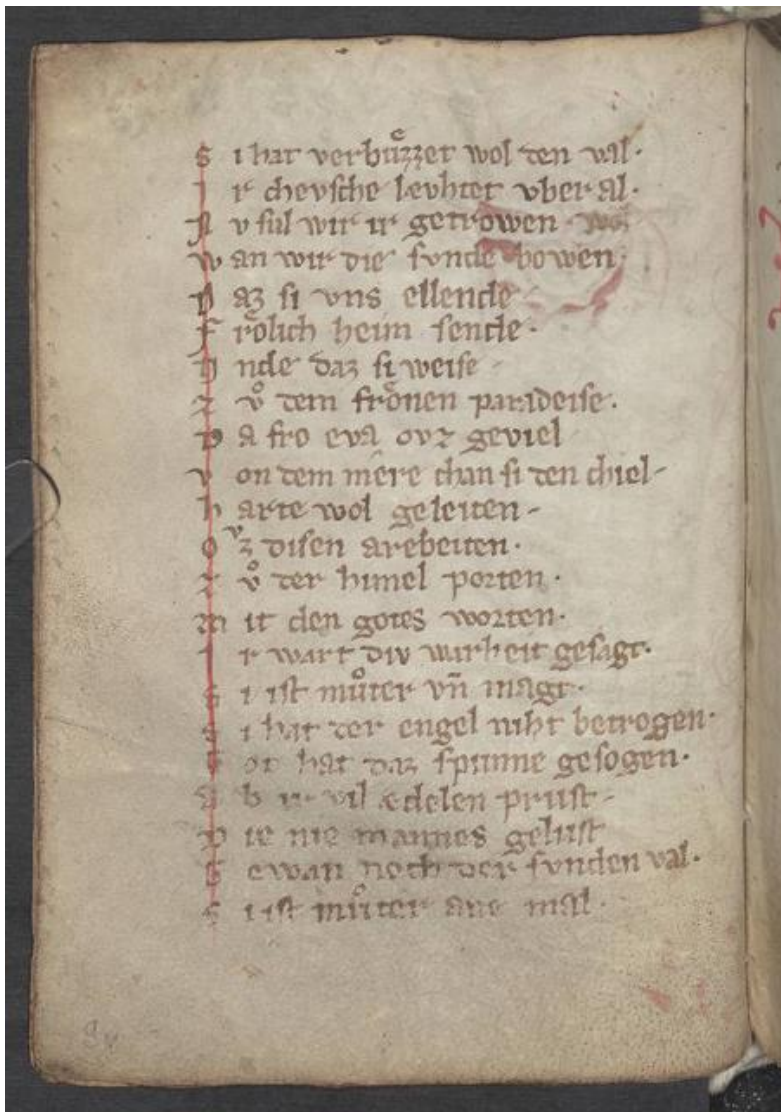
mufen lesen. die gotes kint wellent we

Das er begang mit gûter dat

Daz er begang mit gûter dat

waz er gûtes gelesen hat

waz er gûtes gelesen hat





n wil dar dem iblein manne
 iblein geseht. Vnd dar man sich besee
 vnder dem iblein manne mit vnder pol
He sprach ich wie ich des .viii.
 dute hat vnder geuome das alle die
 die mit hochfuer bekumet sin vnder
 aalter. wie man doch der hochfuer
 lute name vnder. Vnd sprach das
 ee also hochfuerig sy dar sine hee
 mit Carl vnder vnder sin vnder der der
 andes her stier dan ee pol. Vn das
 wie der vnder vnder vnder sollen
 sin die vnder zu hee hant gebu
 Vnd dz wir vnder den freichen bilde
 nime. die des mit vnder vnder
 Vnder die des sechen bilde an
 vnder stier vnder. Vnd vo anden
 vnder alten. **E**n vnder das geseht
 das ma vnder siner vnder einen
 besen hant hant. Vnd wie wie im
 mit vnder vnder. ob er noch vnder
 ist. Vnder die des bilde vnder hee
 vnder vnder vnder vnder vnder
 sprach dann das vnder got am
 maister hat vnder den stier
 wie alle. **I**ch manne de halst
 vnder wie das der tüt der in niege
 sage. Vnder sprach wie iblein das
 stier. Vnder vnder man ee mit sin pol
 in der halst sine vnder vnder sine
 beisse vnder vnder vnder vnder
 vnder. Vnder das das vnder vnder vnder
 vnder das zu iblein hant. Vnder sage
 vnder dem vnder vnder der da
 so lang. Vnder wie man am vnder
 drien hant vnder vnder mag aber am
 besen mit so licht. Vnder wie hant

Ich sol auch das brot essen mit
 Erman bringe die erste recht
 vnder er mit legen pol
 vnder er sich behuten vnder
 vnder halber. In dem vnder
 Er behute sich auch zu der stier
 so er trinke vnder spreche nicht
 die vnder er in dem munde hat nicht
 vnder vnder sich zu den gesellen
 mit dem beche so sie trinken vnder
 E sie in vnder von dem munden
 der von hat sie dar zu gebunden
 vnder trinke vnder den beche seche
 das zimpt hant vnder nicht
 Ein man vor dem gesellen sin
 die esser das ist die hant am
 obe sine do nicht geballe vnder
 so im selber er essen pol
 man pol auch esser alle frist
 mit der hant die vnder ist
 Er der geselle zu der vnder hant
 mit der hant vnder zu hant
 man pol auch das gey vnder
 das man mit esse mit vnder hant
 man pol auch gey nicht
 das man mit dem gemasse nicht
 in die schüssel griffe mit der hant

Experiments – General Methodology

- Using Calamari version 2.1 and mainly its default parameters / best practices
 - Five models per voting ensemble
 - Five augmentations per original line
 - Early stopping after five epochs without improvement
 - ...
- Evaluation
 - Apply models to evaluation data
 - Standardize the output analogously to the training data
 - Calculate the CER using the Calamari evaluation script

Experiments – Training Pipeline and Preliminary Experiments

- Training pipeline to successively refine the models
 - 0: Starting point: LSH-4 (strong MM for material printed between 1450 and 1900)
 - 1: Utilize the entire training set (Gothic + Bastarda)
 - 2: Train two writing-style-specific models by exclusively using Gothic/Bastarda data
- Summary of preliminary experiments
 - Application of in-domain MMs (Gothic → Gothic and Bastarda → Bastarda) yielded best results
 - Out-of-the-box application: very clearly
 - Starting point for finetuning: in-domain > combined > out-of-domain >>> starting from scratch
→ used in-domain MMs for all experiments
 - Calamari predefined network structure *deep3* performed best → used for all experiments (cnn=40:3x3,pool=2x2,cnn=60:3x3,pool=2x2,cnn=120:3x3,lstm=200,lstm=200,lstm=200,dropout=0.5)

Experiments – Simulating an Iterative Training Approach

- Iteratively doubling the amount of training data
 - 2, 4, 8, 16, 32 pages (4 subsumes 2, ...)
 - Starting point: the respective most suitable MM
 - ootb: out-of-the-box result of the MM
- Evaluating on a fixed evaluation set and calculating the CER; averaged over all manuscripts
- **Decent ootb CER** considering the material and MMs
- Training on just two pages **cuts the CER in half**
- CER of 2-3% already completely sufficient for our use-case, many downstream tasks, and efficient further transcription

# pages	CER in %	Impr. in %
ootb	6.22	-
2	3.27	48
4	2.58	21
8	2.17	16
16	1.94	11
32	1.65	15

Example Results After Finetuning on Just Two Pages

uon fo heiliclichem fite. gab im got finen fe
uon fo heiliclichem fite. gab im got finen fe
uon fo heiliclichem fite. gab im got finen fe

gen. himeltov und regen. daz ez michel ge-
gen. himeltov und regen. daz ez michel ge-
gen. himeltov und regen. daz ez michel ge-

nuht gewan. fwaz er buwen began. an fwiv
nuht gewan. fwaz er buwen began. an fwiv
nuht gewan. fwaz er buwen began. an fwiv

fich der herre uerlie. mit heil ez fur fih gie.
fich der herre uerlie. mit heil ez fur fih gie.
fich der herre uerlie. mit heil ez fur fih gie

̄ ergab im fælicliche. er was also vihe riche.
̄ ergab im fælicliche. er was also vihe riche.
̄ ergab im fælicliche. er was also vihe riche.

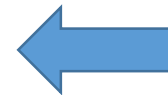
Er erwirbet mit rûm nur daz
Er erwirbet **a**mit rûm **minur** daz
Er erwirbet mit rûm nur daz

Daz ain ieglich man daz gicht
Daz ain ieglich man daz gicht
Daz ain ieglich man daz gicht

Daz sie es tet durch ain bôfwicht
Daz sie es tet durch ain bôfwicht
Daz sie es tet durch ain bôfwicht

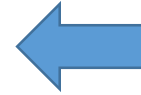
Des ist sie wol ain bôfes wib
Des ist sie wol ain bôfes wib
Des ist sie wol ain bôfes wib

Daz sie gelestert hab **Imrn** lib
Daz sie gelestert hab **Imrn** lib
Daz sie gelestert hab Irn lib



line image
diff view
ground truth

diff view shows **deletions** and **insertions** necessary to turn the prediction into the GT



Example:
model predicted „**m**“,
correct would have been „**rn**“

Experiments – Effect of Mixed Models

- Using MMs as starting points compared to training from scratch
- Unexpectedly vast improvements ...
 - when using small amounts of GT (85% for 2 pages)
 - but also for large training sets (39% for 32 pages)
- Suitable MM not always available or not strong enough but ...
- ... even starting from an out-of-domain MM yields large improvements (>50%) over starting from scratch

# pages	without MM CER in %	with MM CER in %	Impr. in %
2	21.12	3.27	85
4	10.74	2.58	76
8	5.82	2.17	63
16	3.78	1.94	49
32	2.72	1.65	39

Conclusion

- Using a fully open source workflow we ...
 - created a training corpus covering two widely used German medieval writing styles
 - trained two performant and [freely available](#) MMs
- Document-specific finetuning quickly led to very low error rates
 - 2-4 pages of training material already enough to reach our goal of <3% CER in most cases
 - Use of suitable MM as a starting point improved effectiveness and efficiency considerably ... but even a less suited MM is better than nothing
- Future Work
 - Improving our MM as a side-effect of our data collection
 - Opening up new domains by applying the same procedure
 - Automatic quality estimation using Calamari's intrinsic confidence values