

Combining Visual and Linguistic Models for a Robust Recipient Line Recognition in Historical Documents

Martin Mayr, Alex Felker, Mathias Seuret, Andreas Maier, Vincent Christlein

Pattern Recognition Lab, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

Problem Description

- Many (historical) documents available due to mass digitization
- Extraction of meta data is important for historical analysis

Goal

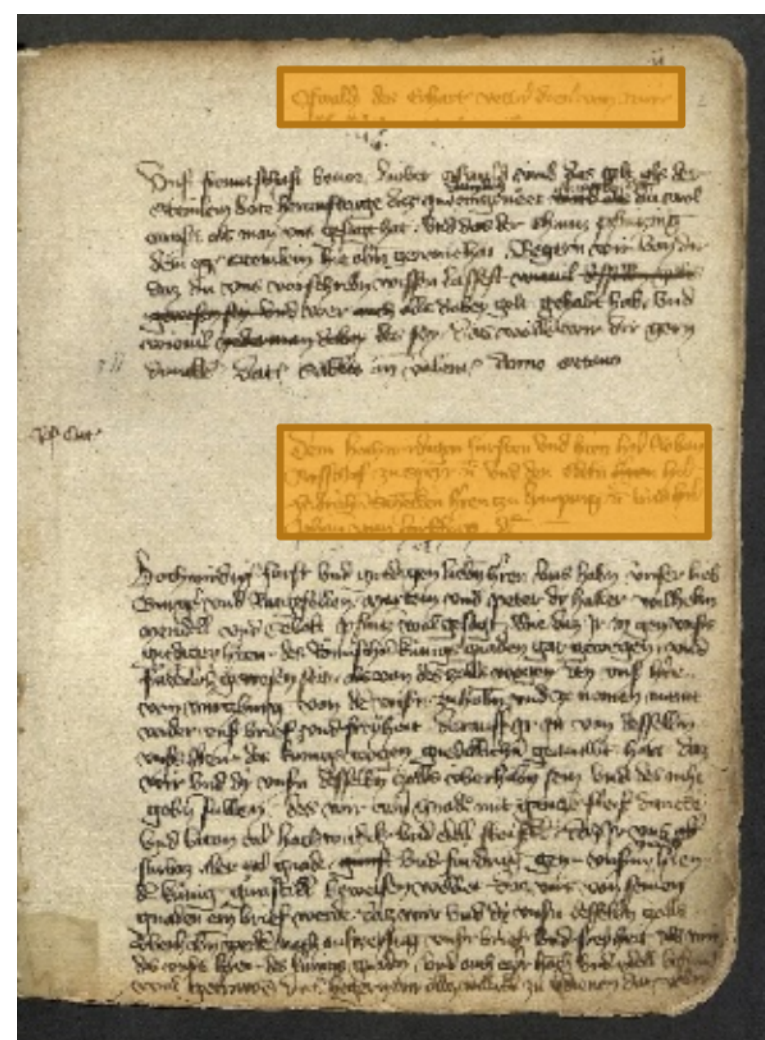
- Find recipient lines in Nuremberg Letterbooks (areas marked in orange)

Problem

- Layout (visual) analysis not robust enough for different books or complicated cases

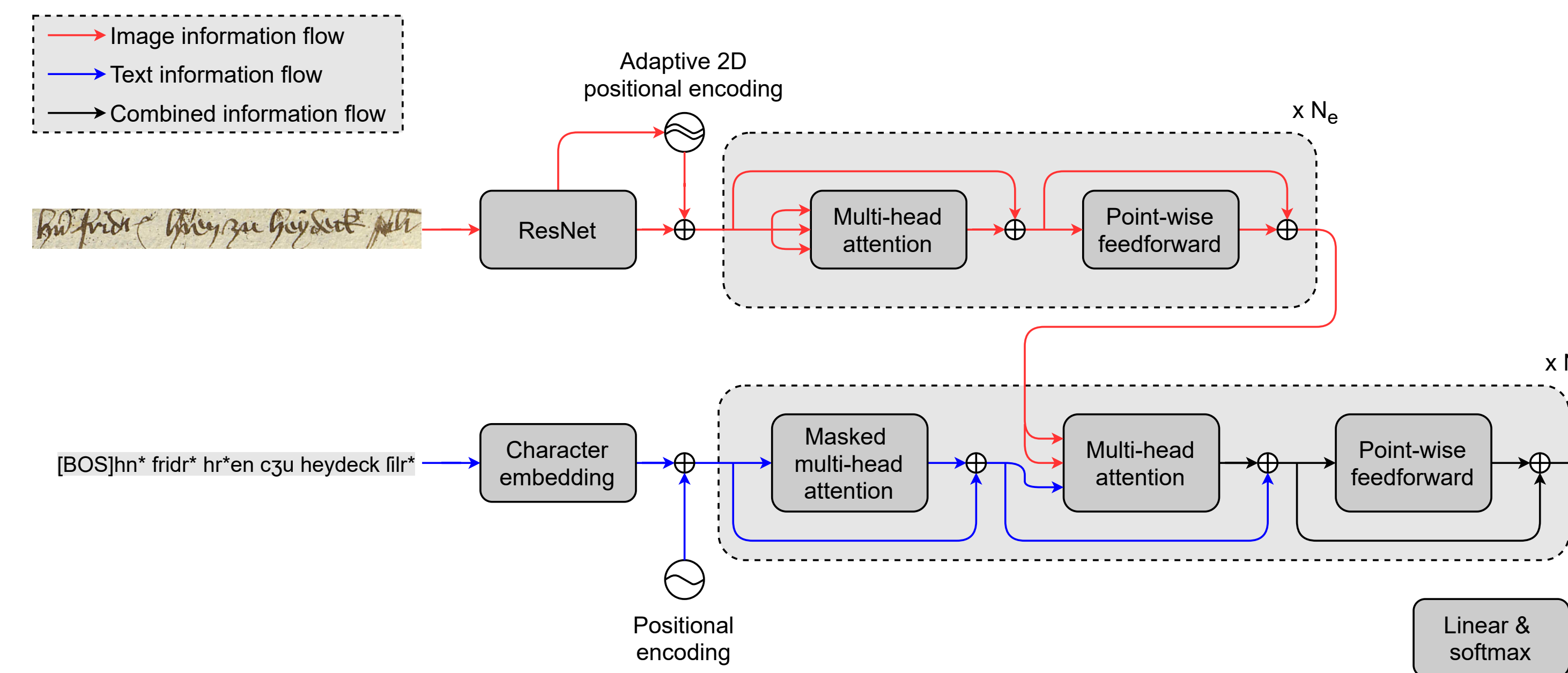
Idea

- Exploit linguistic patterns and combine approaches to improve results



Methodology

Joint Transcriber and Recipient Line Classifier



- Encode classification information into end-of-sequence token of HTR [1]

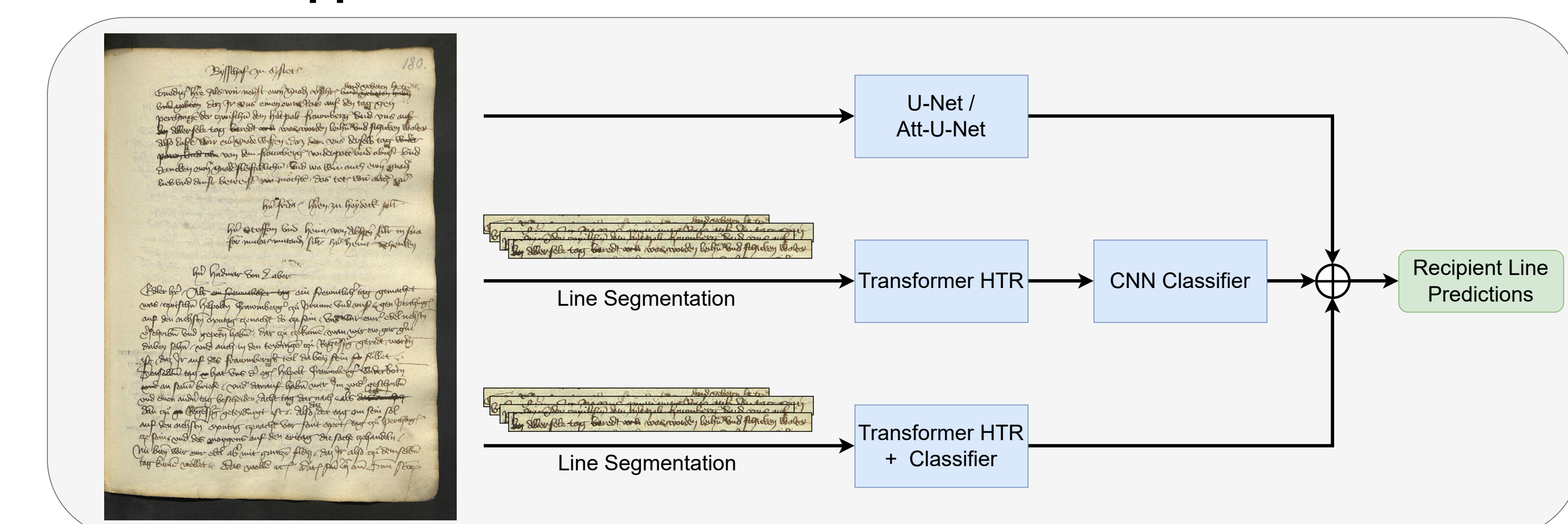
Visual models

- U-Net [2] & Attention U-Net [3]
- No linguistic feedback → only reliant on layout information

Linguistic model

- Transcribe text lines with handwritten text recognition (HTR) model
- Classify recipient lines with n-gram-like feature extraction [4]

Combined Approach



- Combine different modalities from visual and linguistic models
- Each method contributes equally to the combined output
- Note: HTR needs text lines as input and not whole pages

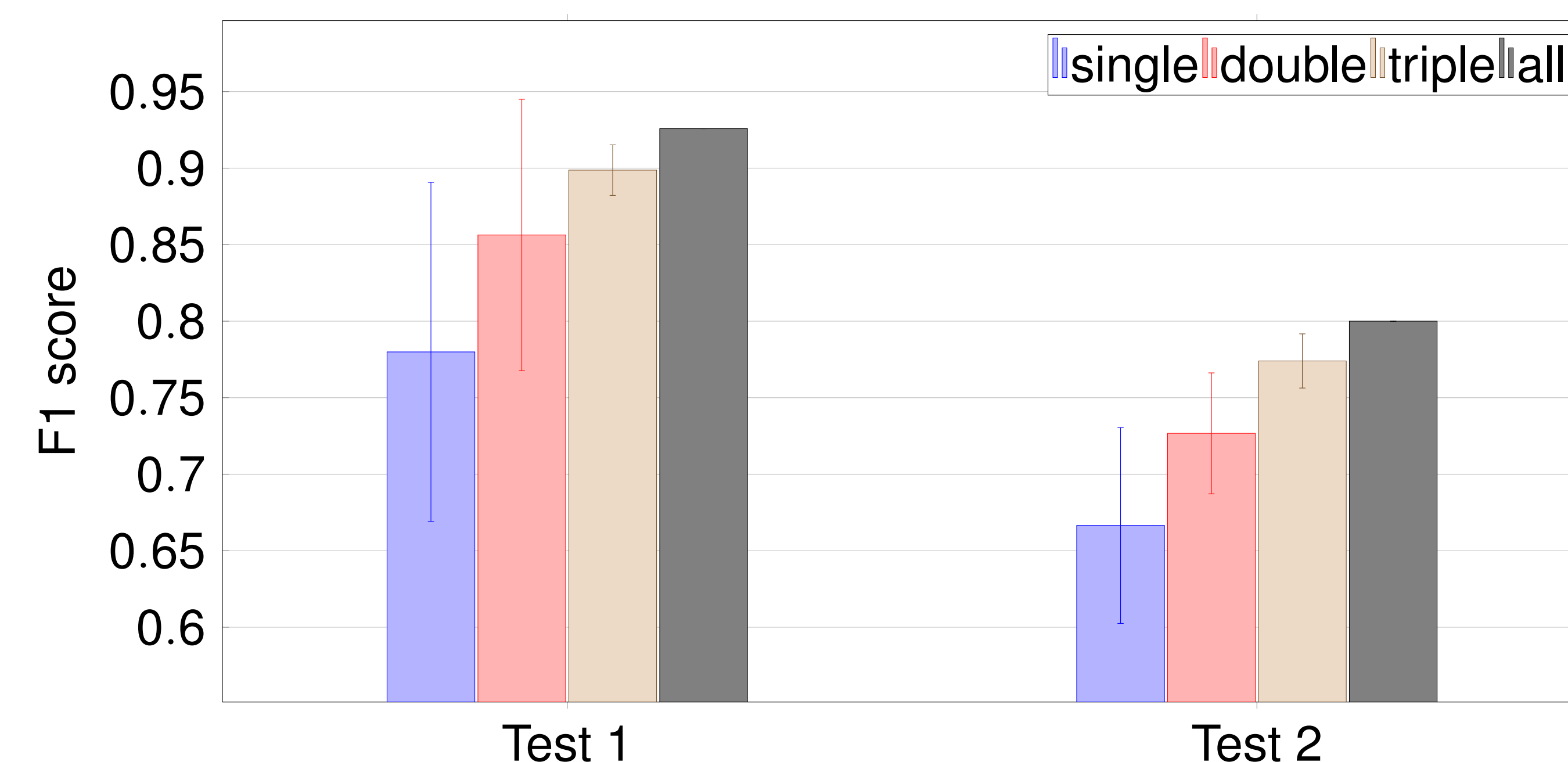
Evaluation

Test cases

- Test 1: same books as in training & validation
- Test 2: unseen book, without curated text line segmentation

	Number of pages			
	Book 2	Book 3	Book 4	Total
Training	–	375	201	576
Validation	–	53	29	82
Test 1	–	102	54	156
Test 2	48	–	–	48
Total	48	530	284	862

Results



Conclusion

- For both test cases the combination of all models works the best (F1 scores: 0.93 (Test 1), 0.80 (Test 2))
- Joint handwriting and recipient line recognition model has a decent performance combined with insights of the decision making
- Project partners are using approach in a semi-automatic way

Outlook

- Improve results with weighting predictions of models according to their performance on validation set
- Evaluate joint model on named entity recognition tasks

References

- A. C. Rouhou, M. Dhiaf, Y. Kessentini, and S. B. Salem, "Transformer-based approach for joint handwriting and named entity recognition in historical document," *Pattern Recognition Letters*, 2021.
- O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), (Cham), pp. 234–241, Springer International Publishing, 2015.
- O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
- M. M. Lopez and J. Kalita, "Deep learning applied to nlp," 2017.

Acknowledgements



Contact



Martin Mayr

Computer Vision
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany

✉ martin.mayr@fau.de