

Bullingers Briefwechsel zugänglich machen: Stand der Handschriftenerkennung

Ströbel, Phillip

pstroebel@cl.uzh.ch
Computer Linguistik, Universität Zürich

Hodel, Tobias

tobias.hodel@unibe.ch
Walter Benjamin Kolleg, Universität Bern, Schweiz

Fischer, Andreas

andreas.fischer@unifr.ch
Institute of Complex Systems (iCoSys), Haute école d'ingénierie et d'architecture de Fribourg, Haute école spécialisée de Suisse occidentale

Scius, Anna

anna.scius-bertrand@hefr.ch
Institute of Complex Systems (iCoSys), Haute école d'ingénierie et d'architecture de Fribourg, Haute école spécialisée de Suisse occidentale

Wolf, Beat

beat.wolf@hefr.ch
Institute of Complex Systems (iCoSys), Haute école d'ingénierie et d'architecture de Fribourg, Haute école spécialisée de Suisse occidentale

Janka, Anna

anna.janka@unibe.ch
Walter Benjamin Kolleg, Universität Bern, Schweiz

Widmer, Jonas

jonas.widmer@unibe.ch
Walter Benjamin Kolleg, Universität Bern, Schweiz

Scheurer, Patricia

patricia.scheurer@uzh.ch
Computer Linguistik, Universität Zürich

Volk, Martin

volk@cl.uzh.ch
Computer Linguistik, Universität Zürich

Automatisierte Handschriftenerkennung fokussierte in der Vergangenheit vorwiegend auf Training und Erkennung einzelner Handschriften, die individuell trainiert wurden. Typischerweise finden sich in den Geisteswis-

senschaften jedoch Datensätze, die von mehreren Händen und häufig auch über längere Zeiträume verschriftlicht wurden. Aussagen zu Erkennalgorithmen müssen entsprechende Voraussetzungen ernst nehmen und nicht nur die Fähigkeit nachweisen, einzelne Hände mit hoher Qualität zu erkennen, sondern auch mit ähnlichen jedoch nicht identischen Handschriften umzugehen.

Ein umfangreicher frühneuzeitlicher Briefwechsel aus dem 16. Jahrhundert, der hauptsächlich in zwei Sprachen (Latein und Frühneuhochdeutsch) vorliegt und mehrere hundert Hände umfasst, ist diesbezüglich ein interessanter Vergleichsfall, um bestehende Plattformen mit neuen Möglichkeiten der *Data Augmentation* und dem Einbezug von Transformer-basierten neuronalen Netzen zu vergleichen. Wir nutzen dafür die im Rahmen des Projekts *Bullinger Digital*¹ erarbeiteten Daten, um Aussagen über Erkennqualität sowie Potenziale in der automatischen Erkennung zu machen. Für die Evaluation verwenden wir zwei spezifisch erstellte Testsets. Diese widerspiegeln die Tatsache, dass von einigen (wenigen) Personen eine Vielzahl von Briefen und von diversen Personen nur wenige Schreiben im Datensatz vorhanden sind (Abbildung 1). Die beiden Testsets sind zwar ähnlich groß (**Viel-** (1'235 Zeilen) als auch **Wenigschreiber** (1'013 Zeilen)), leisten aber Aussagen zu sehr unterschiedlichen Größenordnungen, da von den Vielschreibern eine weit umfangreichere Masse erkannt werden wird, aber auch mehr Material zum Training von Modellen zur Verfügung steht. Bei der Evaluation können wir aufgrund der Aufteilung aber präzisere Voraussagen machen, welche Datenmassen mit welcher Qualität erkannt werden. Weiter lässt sich abschätzen, welche Erkennungsform sich für welche (Trainings-)Datenmenge eignet.

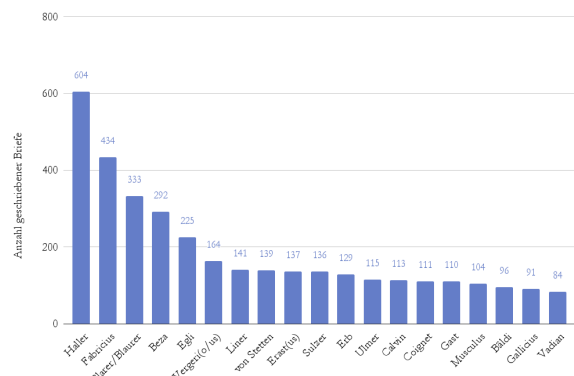


Abbildung 1. Verteilung der Top-20-Autoren, welche Briefe an Bullinger schrieben (Anzahl Briefe auf y-Achse).

Im Folgenden testen wir drei Ansätze, die Aufschlüsse zum Stand und möglichen Entwicklungen im Bereich der Handschriftenerkennung versprechen, am Korpus:

1. *Transkribus* mit seinen zwei Engines (HTR+ (Strauss et al. 2018) und PyLaia (Puigcerver 2017)) testet eine etablierte Methode am Datensatz.
2. *Data Augmentation* soll die Erkennung mit der State-of-the-Art-Engine HTR-Flor (de Sousa Neto et al. 2020) verbessern.

3. Neue Transformer-basierte Modelle (Li et al. 2021) sollen auf ihre Tauglichkeit für historische Daten geprüft werden.

Seit 1974 erscheinen in regelmäßigen Abständen Bände der Heinrich-Bullinger-Briefwechsel-Edition, erarbeitet am Institut für Schweizerische Reformationsgeschichte (IRG) der Universität Zürich. Der letzte Band stammt von 2022 und weitere sind aktuell in Arbeit (Bullinger 2022). Die Edition bedient diverse Nutzer*innengruppen, insbesondere Theolog*innen und die Geschichtswissenschaften, was sich etwa an den Editionsgrundsätzen (Bullinger 1973) zeigt, z. B. stillschweigende Auflösung von Abkürzungen, Identifikation von Personen und Orten etc.

Um die Vielzahl der nicht transkribierten Briefe schneller maschinell zu verarbeiten, arbeiten wir im Rahmen des Projekts *Bullinger Digital* an der automatisierten Aufbereitung der Dokumente. Die Texte der bereits edierten Briefe sowie der am IRG provisorisch transkribierten Briefe haben wir mittels des *Text-to-Image-Verfahrens* (Leifert, Labahn, and Sánchez 2020) mit den Bildvorlagen automatisiert aligniert. Damit steht nun ein Korpus von 1'297'908 Token für Training und Validierung zur Verfügung.²

Bei der Zuordnung der Zeilen wurde mit einem *Threshold* agiert, der dafür sorgte, dass nur transkribierte Textteile zugeordnet wurden, bei denen eine relativ hohe Wahrscheinlichkeit der Übereinstimmung errechnet wurde. Dadurch wurden zwar ca. 25-30% aller Zeilen nicht zugeordnet, Fehler in der Layouterkennung und nicht identifizierte Streichungen und Ähnliches führten aber nicht zu falschen Zuordnungen, die wiederum negativen Einfluss auf das Training von Texterkennungsmodellen hätte.

Texterkennung mit etabliertem Framework und Plattform: Transkribus

2015 wurde die Plattform *Transkribus* gelauncht und danach von 2016 bis 2019 *Deep-Learning*-basierte Erweiterungen, insbesondere Layout- und Texterkennungsgines (Muehlberger et al. 2019) implementiert. Die Plattform wird von einer Kooperative betrieben und für diverse Zwecke in der Forschung und durch Erinnerungsinstitutionen eingesetzt. Training von Handschriftenmodellen und Erkennung neuer Seiten erfolgen über ein GUI.

Basierend auf den oben erwähnten Trainings- und Testdaten erzeugten wir mehrere Handschriftenmodelle, dabei testeten wir jeweils beide verfügbaren Engines (HTR+ und PyLaia), um Unterschiede in der Qualität aufzuzeigen. Wir übernahmen die Layouterkennung unverändert.

Die in Transkribus trainierten Modelle (Tabelle 1) zeigen gute Ergebnisse. Gerade für Vielschreiber wird trotz vieler unterschiedlicher Hände insbesondere mit HTR+ eine Erkennung unter 8% *Character Error Rate* erreicht.³ Die Erkennung der Wenigschreiber ist dagegen etwas fehlerbehafteter. Alle Modelle wurden "austrainiert", ohne dass

ein problematisches *Overfitting* beobachtet wurde (Hodel 2020).

Eine signifikante Verbesserung erreichten wir durch die Nutzung von grossen Modellen als *Basemodels*. Mit einem vortrainierten Modell basierend auf 5'820'990 Wörtern (im Sinne von *Tokens*) in lateinischer Schrift (keine Kurrentschrift), kann die Erkennungsqualität weiter verbessert werden.

Tabelle 1. Resultate der sprachunabhängig trainierten Modelle in Transkribus mit HTR+ und PyLaia. HTR+: 500 Epochen, PyLaia: 250 Epochen.

		CER			
		Vielschreiber	Wenigschreiber		
Trainingsdaten	base model	HTR+	PyLaia	HTR+	PyLaia
multilingual	-	7.26	9.9	10.06	12.7
	Latin	6.79	-	9.51	-

Data Augmentation zur Erweiterung des Trainingsmaterials

Um für einen gegebenen Schreibstil ein zuverlässiges Erkennungssystem zu trainieren, braucht es eine große Anzahl annotierter Textzeilen. Es ist deshalb schwierig, seltene Hände automatisch zu transkribieren, da der entsprechende Schreibstil nicht oder nur ungenügend in den Trainingsdaten repräsentiert ist. *Data Augmentation* kann verwendet werden, um die Trainingsdaten automatisch mit weiteren Beispielen anzureichern und so den Lernprozess zu unterstützen. In unserer Arbeit verfolgen wir dazu einen vielversprechenden Ansatz: Wir lernen die Schreibstile von seltenen Händen mittels *Generative Adversarial Networks (GANs)*, um anschließend beliebige Texte zu synthetisieren und den Trainingsdaten als zusätzliche Lernbeispiele hinzuzufügen.

Für die Synthetisierung wird *lineGen* (Davis et al. 2020) eingesetzt, ein kürzlich vorgeschlagenes GAN-Netzwerk, welches auf ganzen Textzeilen arbeitet und drei Zielfunktionen integriert: den *Adversarial Loss* (Unterscheidung zwischen echten und synthetischen Bildern), den *Perceptual Loss* (visuelle Qualität der generierten Bilder) und den *CTC Loss* (Qualität der automatischen Transkription). Das trainierte lineGen-System erlaubt es, einen beliebigen Text mit dem gelernten Stil zu synthetisieren, d. h. aus dem Text ein Textzeilenbild zu generieren. Abbildung 2 zeigt Beispiele von generierten Textzeilenbildern, wenn lineGen auf rund 17'000 Textzeilenbildern unterschiedlich lange und unter Verwendung der Standard-Hyperparameter mit verschiedenen Schreibstilen trainiert wird und im Anschluss ein englischer Text synthetisiert wird. Das GAN konvergiert auf einen Schreibstil, der ähnlich leserlich ist wie echte Beispiele aus den Trainingsdaten.⁴

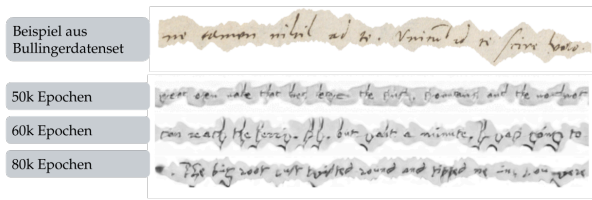


Abbildung 2. Beispiele von synthetischen Textzeilenbildern basierend auf lineGen über unterschiedliche Anzahl Epochen.

Der Einfluss der synthetischen Lernbeispiele auf die Erkennungsrate wurde in zwei Szenarien untersucht, welche eine kleine Trainingsmenge vorsehen, wie es für Wenigschreiber typisch ist: In Szenario A gehen wir von 1'000 Trainingszeilen aus und in Szenario B von 200 Trainingszeilen. Als Erkennungssystem wird HTR-Flor (de Sousa Neto et al. 2020) eingesetzt, eines der besten Systeme nach aktuellem Stand der Technik, und für die Auswertung werden rund 1'000 zufällig ausgewählte Textzeilen als Testdaten verwendet. Abbildung 3 zeigt den Trainingsverlauf für Szenario A unter Verwendung der Standard-Hyperparameter für HTR-Flor. Die CER auf den Testdaten erreicht 26.8% für das Training mit 1'000 echten Textzeilen. Wenn mit 1'000 synthetischen Textzeilen trainiert wird, scheitert das Lernen. Eine mögliche Interpretation ist, dass die synthetische Schrift nicht genügend natürliche Variation beinhaltet, welche fürs Trainieren nötig ist. Wenn aber die 1'000 echten mit den 1'000 synthetischen Textzeilen kombiniert werden, kann die CER signifikant auf 24.1% verbessert werden. Abbildung 4 zeigt den Trainingsverlauf für Szenario B. Hier scheitert das Lernen sowohl mit 200 echten als auch mit 200 synthetischen Textzeilen. Hingegen führt die Kombination von echten und synthetischen Trainingszeilen erneut zu einer signifikant verbesserten CER von 47.7%.

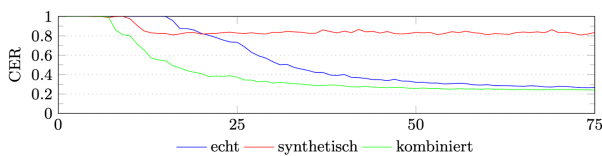


Abbildung 3. Training von HTR-Flor während 75 Epochen im Szenario A (1'000 Trainingszeilen).

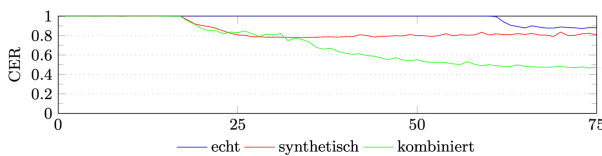


Abbildung 4. Training von HTR-Flor während 75 Epochen im Szenario B (200 Trainingszeilen).

Diese initialen experimentellen Resultate sind vielversprechend und legen die Vermutung nahe, dass GAN-basierte synthetische Lernbeispiele dabei helfen können, die Erkennung für Wenigschreiber zu verbessern. In einem nächsten Schritt planen wir umfangreichere Experimente, welche darauf abzielen, den Stil von Wenigschreibern zu lernen oder den Stil von Gruppen ähnlicher

Schriftbilder. Eine wichtige Fragestellung dabei wird sein, wie viele Lernbeispiele nötig sind, um einen Stil zuverlässig lernen zu können. Anschließend planen wir, ein *Transfer Learning* durchzuführen, ausgehend von einem gut trainierten Grundsystem für Vielschreiber, welches mit Hilfe der synthetischen Lernbeispiele an die Wenigschreiber angepasst wird.

Nutzung einer Transformer-basierten Architektur

Transformer-basierte Architekturen (Vaswani et al. 2017) haben das Feld der natürlichen Sprachverarbeitung in Sachen Sprachmodellierung revolutioniert und zu einem Paradigmenwechsel beim Trainieren von Systemen für unterschiedlichste Anwendungen geführt. Transformer-Modelle wie BERT (Devlin et al. 2018) und deren Weiterentwicklungen, welche auf großen Mengen an Sprachdaten trainiert wurden, sind als starke Transfer-Lerner (Ruder et al. 2019) bekannt und erlauben einen vielfältigen Einsatz beim Fine-Tuning für spezifische Aufgaben (*Natural Language Understanding*, Fragenbeantwortung, Wortarten-Klassifikation). Transformer können auch für Bildverarbeitung genutzt werden (Dosovitskiy et al. 2021; Touvron et al. 2021), was die Entwicklung von BERT-ähnlichen und auf großen Bildmengen vortrainierten Modellen zur Folge hatte (Bao, Dong, and Wei 2021).

Für unser Projekt nutzen wir als Grundlage *TrOCR* (Li et al. 2021), welchem ein *Encoder-Decoder*-Struktur zugrunde liegt. Bildseitig beruht der Encoder auf der BEiT-Architektur (Bao, Dong, and Wei 2021), welche für die *Feature*-Extraktion aus den Bildern verantwortlich ist, während der *Decoder* die Bildinformation mit Hilfe eines RoBERTa-Sprachmodells (Liu et al. 2019) in eine *Subword*-Folge "übersetzt". Li et al. benutzten in zwei *Pre-Training*-Phasen über 700 Millionen an synthetisch erzeugten und echten Textzeilenbildern mit dem dazugehörigen Text in englischer Sprache, die danach mit dem IAM-Datensatz (Marti and Bunke 2002) einem *finetuning* unterzogen wurden. Die CER von 2.89% liegt nur 0.14 Prozentpunkte hinter einem klassischen Ansatz (Diaz et al. 2021).

Im Gegensatz zu den beiden anderen Projektteilen wurde für die *TrOCR*-Anwendung die Zeilen nach Sprachen unterschieden. Diese Sprachverteilung und einige weitere Kennzahlen können Tabelle 2 entnommen werden.

Tabelle 2. Kennzahlen der extrahierten Zeilen aus dem Material der Bullinger-Briefe.

	# Zeilen	%	# Worte	# Wörter / Zeile	# Buchstaben	# Buchstaben / Zeile
Latein	134'236	81.02	1'073'106	7.99	7'314'648	54.49
FNHD	31'437	18.98	253'372	8.06	1'547'725	49.23
Total	165'673		1'326'478	8.01	8'862'373	53.49

Wir untersuchen die Eignung von *TrOCR* für die Texterkennung auf historischen Daten, indem wir die Anzahl Epochen, die für das Fine-Tuning aufgewendet werden, variieren und multi- wie auch monolinguale Modelle trainieren. Alle Modelle führen während eines Drittels der Epochenzahl ein *Warm-Up* durch.⁵ Für die Evaluation ha-

ben wir auch die Testsets nach Sprachen aufgeteilt. Tabelle 3 fasst die Ergebnisse zusammen.

Tabelle 3. Resultate der Modelle trainiert auf verschiedenen Sprachzusammenstellungen und Anzahl Epochen (fett = beste Performance in einer Spalte).

		CER		FNHD	multiling.	Latein	FNHD
		Vielschreiber	Wenigschreiber				
Trainingsdaten	# Epochen	multiling.	Latein	FNHD	multiling.	Latein	FNHD
multilingual	1	9.53	9.49	9.7	11.58	11.07	12.79
	2	8.38	8.51	7.71	10.46	10.18	11.11
	3	7.81	8.07	6.5	9.95	9.61	10.74
	4	7.21	7.61	5.27	9.68	9.31	10.55
	5	7.31	7.85	5.25	9.69	9.54	10.25
	6	7.41	7.89	5.09	9.61	9.39	10.11
Latein	5	11.09	7.99	26.21	16.1	9.69	31.21
FNHD	5	28.41	33.02	6.06	27.89	34.83	11.43

Aufgrund der Resultate entschieden wir uns, die monolingualen Modelle auf 5 Epochen zu trainieren. Auf die frühneuhochdeutschen Zeilen der Vielschreiber angewandt liefert das monolinguale Modell mit einer CER von 6.06% entgegen der Erwartungen nicht das beste Resultat. Dieses erreichte ein multilinguales Modell (trainiert über 6 Epochen) mit einer CER von 5.09%, also knapp einem Prozentpunkt Unterschied. Die größere Menge an Bilddaten beeinflusst während des Fine-Tunings merklich die Performanz.

Unsere Experimente zeigen, dass Transformer-basierte HTR für historische Daten CERs in akzeptablen Bereichen liefert, ohne dass TrOCR bis zu unserem Fine-Latein oder FNHD gesehen hätte. Trotz der im Vergleich zum Pre-Training kleinen lateinischen und frühneuhochdeutschen Textmenge lernt TrOCR die Dekodierung von neuen Sprachen zuverlässig. In Zeilen, in welchen die Sprache hingegen ändert (Code-Switching), was im Bullinger-Briefwechsel relativ häufig passiert (Volk et al. 2022), reagiert TrOCR zu spät für den Sprachwechsel (Abbildung 5). Solche Phänomene werden wir mit flexibleren Modellen abzufangen versuchen.

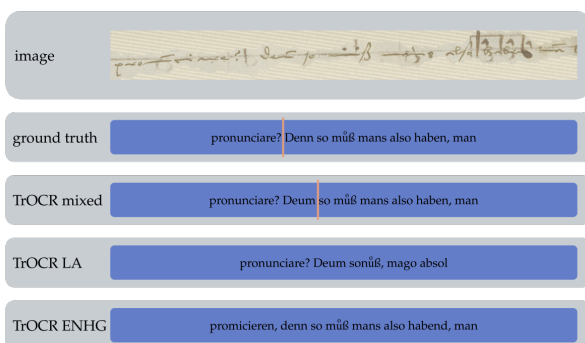


Abbildung 5. Performanz verschiedener TrOCR-Modelle auf einer Zeile, in welcher Sprachwechsel vorliegt. Der Trennstrich markiert die Sprachgrenze, die im multilingualen Modell zu spät und in monolingualen gar nicht erkannt wird.

Schlüsse

Die Arbeit mit dem Bullinger-Korpus ist aufschlussreich in unterschiedlicher Hinsicht. Zentral für diese Arbeit sind drei *Schlussfolgerungen*:

1. Wir stellen eine Harmonisierung unterschiedlicher (*Deep-Learning*-basierter) Systeme mit Bezug zur Erkennqualität von Handschriften fest. Unabhängig davon, ob etablierte Plattformen oder neue Systeme wie TrOCR genutzt werden.
2. Der Gebrauch von *Data-Augmentation*-Techniken verspricht einen Gewinn, der aktuell noch weiter auszuloten ist, im Grundsatz aber schon zu Verbesserungen führt.
3. Aktuell haben etablierte Plattformen den Vorteil, dass grosse *Basemodels* genutzt werden können, die noch leichte Vorteile mitbringen, bei grossen Projekten wie Bullinger Digital aber noch anhand der Fehlermuster ausgelotet werden müssen.

Fußnoten

1. <https://www.bullinger-digital.ch/>.
2. Stand 1.6.2022.
3. Character Error Rate wird mit CER abgekürzt und steht für Zeichenfehlerrate.
4. Für den Aufbau des Experiments: Spoto et al. 2022.
5. Warm-up bedeutet, dass zu Beginn des Trainings mit kleinen Lernraten gerechnet wird, um das Modell langsam an die neuen Daten zu gewöhnen.

Bibliographie

- Bullinger, Heinrich.** 1973. *Briefe der Jahre 1524-1531*. Edited by Ulrich Gäbler and Fritz Büsser. Vol. 1. Heinrich Bullinger Werke. Zweite Abteilung: Briefwechsel. Zürich: Theologischer Verlag.
- . 2022. *Briefe von April bis Dezember 1547: Anhang: Neue Briefe aus den Jahren 1523 bis 1546*. Edited by Reinhard Bodenmann, Yvonne Häfner, and Judith Steiniger. 1st ed. Vol. 20. Heinrich Bullinger Werke. Zweite Abteilung: Briefwechsel. Zürich: Theologischer Verlag Zürich.
- Davis, Brian, Bryan Morse, Brian Price, Chris Tensmeyer, Curtis Wington, and Rajiv Jain.** 2020. "Text and Style Conditioned GAN for the Generation of Offline-Handwriting Lines." In *The 31st British Machine Vision (Virtual) Conference 2020*. Bath. https://www.bmvc2020-conference.com/conference/papers/paper_0815.html.
- Diaz, Daniel Hernandez, Siyang Qin, Reeve Ingle, Yasuhisa Fujii, and Alessandro Bissacco.** 2021. "Rethinking Text Line Recognition Models." arXiv. <http://arxiv.org/abs/2104.07787>.
- Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, et al.** 2021. "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv. <http://arxiv.org/abs/2010.11929>.

Hodel, Tobias. 2020. "Best-Practices Zur Erkennung Alter Drucke Und Handschriften – Die Nutzung von Transkribus Large- Und Small-Scale." In *DHd 2020. Spielräume Digital Humanities Zwischen Modellierung Und Interpretation*, edited by Christof Schöch, 84–87. Paderborn. <https://doi.org/10.5281/zenodo.3666689>.

Leifert, Gundram, Roger Labahn, and Joan Andreu Sánchez. 2020. "Two Semi-Supervised Training Approaches for Automated Text Recognition." In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 145–50. <https://doi.org/10.1109/ICFHR2020.2020.00036>.

Li, Minghao, Tengchao Lv, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, and Furu Wei. 2021. "TrOCR: Transformer-Based Optical Character Recognition with Pre-Trained Models." arXiv. <http://arxiv.org/abs/2109.10282>.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv. <https://doi.org/10.48550/arXiv.1907.11692>.

Marti, U.-V., and H. Bunke. 2002. "The IAM-Database: An English Sentence Database for Offline Handwriting Recognition." *International Journal on Document Analysis and Recognition* 5 (1): 39–46. <https://doi.org/10.1007/s100320200071>.

Puigcerver, Joan. 2017. "Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?" In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 01:67–72. <https://doi.org/10.1109/ICDAR.2017.20>.

Ruder, Sebastian, Matthew E. Peters, Swabha Swayamdipta, and Thomas Wolf. 2019. "Transfer Learning in Natural Language Processing." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, 15–18. Minneapolis, Minnesota: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-5004>.

Sousa Neto, Arthur Flor de, Byron Leite Dantas Bezerra, Alejandro Héctor Toselli, and Estanislau Baptista Lima. 2020. "HTR-Flor: A Deep Learning System for Offline Handwritten Text Recognition." In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 54–61. <https://doi.org/10.1109/SIBGRAPI51738.2020.00016>.

Spoto, Martin, Beat Wolf, Andreas Fischer, and Anna Scius-Bertrand. 2022. "Improving Handwriting Recognition for Historical Documents Using Synthetic Text Lines." In *Proceedings 20th Conf. of the International Graphonomics Society (IGS)*. Las Palmas de Gran Canaria.

Strauss, Tobias, Max Weidemann, Johannes Michael, Gundram Leifert, Tobias Grüning, and Roger Labahn. 2018. "System Description of CITLab's Recognition & Retrieval Engine for ICDAR2017 Competition on Information Extraction in Historical Handwritten Records." *CoRR* abs/1804.09943. <http://arxiv.org/abs/1804.09943>.

Touvron, Hugo, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Herve Jegou. 2021. "Training Data-Efficient Image Transformers & Distillation through Attention." In *Proceedings of the 38th International Conference on Machine Learning*, 10347–57. PMLR. <https://proceedings.mlr.press/v139/touvron21a.html>.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, 30:5998–6008. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>.

Volk, Martin, Lukas Fischer, Patricia Scheurer, Raphael Schwitter, Phillip Ströbel, Benjamin Suter. 2022. "Nunc Profana Tractemus. Detecting Code-Switching in a Large Corpus of 16th Century Letters." In *Proceedings of LREC 2022*. Marseille.