Daniel Schneider, Nikolaus Korfhage, Markus Mühling, Peter Lüttig, Bernd Freisleben

DeepTab: Automatische Transkription von Orgeltabulaturen in die heutige Notenschrift mittels tiefer neuronaler Netze

Zusammenfassung

Die Orgeltabulaturschrift ist eine Notenschrift, die von der Mitte des 14. bis zum Anfang des 18. Jahrhunderts eingesetzt wurde. Sie unterscheidet sich in Aufbau und Form deutlich von der heutigen Notenschrift, so dass für musikwissenschaftliche Analysen von Orgeltabulaturwerken üblicherweise eine Übertragung in die heutige Notenschrift vorgenommen wird. Eine manuelle Übertragung ist jedoch ein sehr zeitaufwändiger und fehleranfälliger Prozess. Daher existieren zu vielen ausschließlich in Orgeltabulaturschrift überlieferten Musikstücken bislang keine Übertragungen in die heutige Notenschrift.

In diesem Beitrag präsentieren wir *DeepTab*, ein Software-Werkzeug, das eingescannte Orgeltabulaturseiten automatisiert in die heutige Notenschrift überträgt. Die technische Grundlage von *DeepTab* ist ein künstliches neuronales Netz, das mit einer Kombination von realen, manuell annotierten Zeilen aus zwei Orgeltabulaturbüchern sowie künstlich erzeugten Bildern zufallsgenerierter Orgeltabulaturzeilen trainiert wurde.

DeepTab liefert eine einheitliche Transkription von Orgeltabulaturen ohne Fehlerkorrekturen oder weitere Interpretationen, bleibt also so nahe am Original wie möglich, was insbesondere für die musikwissenschaftliche Forschung unabdingbar ist.

1. Einleitung

Die Untersuchung historischer Musiknotationen ist ein wichtiges Forschungsthema in der Musikwissenschaft. In vielen Fällen ist für eine musikwissenschaftliche Analyse von Stücken in einer alten Notation eine Transkription in die heutige Notenschrift erforderlich. Auch eine Aufbereitung in digitalen Notenformaten wie LilyPond[1] oder MusicXML[2] kann oftmals hilfreich sein. Eine Übertragung in die heutige Notenschrift ermöglicht es, die Stücke einem größerem Personenkreis zu Forschungs- oder Aufführungszwecken zugänglich zu machen. Manuelle Transkriptionen sind jedoch häufig sehr zeitaufwändig und fehleranfällig. Eine solche alte Art der Musiknotation ist die sogenannte „Neue Deutsche Orgeltabulatur“, die von der Mitte des 16. bis zum Anfang des 18. Jahrhunderts eingesetzt wurde und für Musikwissenschaftler/-innen eine wichtige Quelle des Wissens über die Musik der Renaissance darstellt. In mehreren Archiven findet sich eine Vielzahl von in Orgeltabulaturnotation geschriebenen Musikstücken,

1 http://www.lilypond.org.
2 https://www.musicxml.com.

von denen einige bisher weder digitalisiert, noch in die heutige Notenschrift übertragen wurden.[3]

In diesem Beitrag präsentieren wir ein Software-Werkzeug, genannt *DeepTab*, das analog zur automatisierten Texterkennung in gescannten Dokumenten, der sogenannten Optical Character Recognition (OCR), eine automatisierte Erkennung von Orgeltabulaturzeichen ermöglicht.[4] *DeepTab* analysiert gescannte Orgeltabulaturseiten und überträgt sie automatisiert in die heutige Notenschrift.

Die Transkription läuft in einem mehrstufigen Prozess ab. Zunächst wird jedes gescannte Eingabebild in die einzelnen darauf abgebildeten Tabulatursysteme segmentiert. Auf den entstehenden Teilbildern wird mithilfe eines künstlichen neuronalen Netzes (NN) eine Erkennung der Tabulaturzeichen durchgeführt. Die Ergebnisse werden in das Format des frei verfügbaren Notensatzprogramms LilyPond[5] überführt, das zur Erzeugung einer grafischen Ausgabe in heutiger Notation verwendet wird. Ein Beispiel für eine solche automatische Transkription ist in ABBILDUNG 1 dargestellt.

Um das neuronale Netz auf die Aufgabe der Tabulaturzeichenerkennung zu trainieren, werden annotierte Orgeltabulaturbilder benötigt, also Bilder mit einer Beschriftung der darauf abgebildeten Zeichen. Zu diesem Zweck wurden Bilder einzelner Tabulatursysteme aus zwei vollständig als Scan verfügbaren Orgeltabulaturbüchern extrahiert und mit den darauf dargestellten Zeichensequenzen annotiert. Um die Menge und Vielfalt der Trainingsdaten weiter zu erhöhen, wurde ein Datengenerator entwickelt, der ausgehend von Bildern einzelner Tabulaturzeichen künstliche Bilder von Orgeltabulaturzeilen zufällig generiert. Das Training erfolgte mit einer Kombination dieser künstlich generierten Daten und aus den Büchern extrahierten Realdaten.

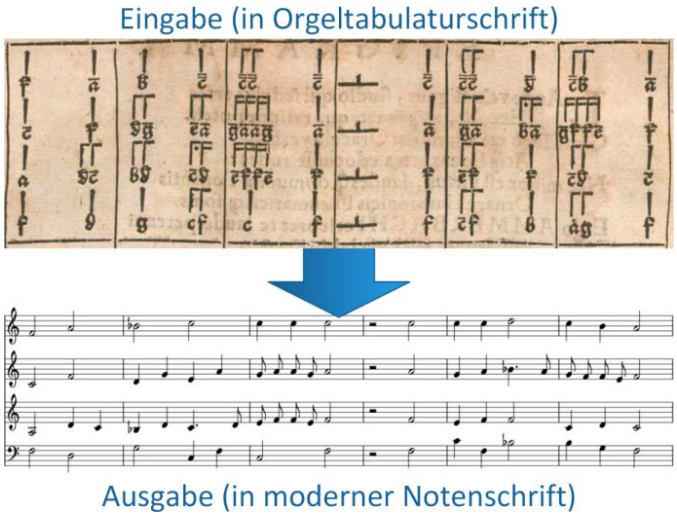

ABBILDUNG 1: TRANSKRIPTION EINES VIERSTIMMIGEN TABULATUR-NOTENSYSTEMS AUS DEM *ORGEL- ODER INSTRUMENT TABULATURBUCH* VON ELIAS NIKOLAUS AMMERBACH[6] IN DIE HEUTIGE NOTENSCHRIFT.

3 Marko Motnik, „Deutsche Tabulatur: Gebreuchlich oder verdrießlich?“, in: *Musicological Annual* 47.2 (2011), S. 125–137, ISSN: 0580373X; Elżbieta Wojnowska, *Thematic Catalogue of 17th-Century Organ Tablatures from the Liegnitz Bibliotheca Rudolphina*, Warschau 2016.

4 Daniel Schneider et al., „Automatic transcription of organ tablature music notation with deep neural networks“, in: *Transactions of the International Society for Music Information Retrieval* 4.1 (2021), S. 14–28. DOI: https://doi.org/10.5334/tismir.77.

5 Siehe Anm. 1.

6 Elias Nikolaus Ammerbach, *Orgel oder Instrument Tabulaturbuch*, Nürnberg 1583. https://imslp.org/wiki/Special:ReverseLookup/286314 [letzter Zugriff am 16.02.2022].

Der Rest dieses Beitrags ist wie folgt gegliedert. In Abschnitt 2 folgt ein Überblick über die Entstehung und den Aufbau der Orgeltabulaturschrift und die damit einhergehenden Anforderungen an eine OCR-Anwendung. Abschnitt 3 erläutert Grundlagen zu künstlichen neuronalen Netzen. Darauf aufbauend stellt Abschnitt 4 *DeepTab* genauer vor. Abschnitt 5 geht auf die zum Training und zur Evaluation des neuronalen Netzes eingesetzten Datensätze ein und beschreibt den Datengenerator für künstliche Orgeltabulaturbilder. Die Ergebnisse einer Evaluation von *DeepTab* und eine genauere Betrachtung der Analysefehler erfolgt in Abschnitt 6. Die Arbeit schließt mit einer Zusammenfassung und einem Ausblick auf zukünftige Forschungsarbeiten.

2. Orgeltabulaturschrift

2.1. Entstehung von Tabulaturschriften

Tabulaturschriften sind eine alternative Form der Notation von Musikstücken, die sich von der heutigen Notenschrift mit fünf Notenlinien deutlich unterscheiden. Der Begriff Tabulatur leitet sich vom lateinischen Wort „tabula“ (Tafel, Blatt) ab, der als Bezeichnung für das Blatt, auf dem ein Musikstück niedergeschrieben wurde, diente.[7]

Die ersten Formen der Tabulaturschrift entstanden im Mittelalter, als sich die Instrumentalmusik mehr und mehr von ihren volkstümlichen Ursprüngen hin zu einer Kunstform entwickelte, womit auch der Bedarf für eine schriftliche Aufzeichnung der Musik wuchs. Tabulaturen bildeten eine Alternative zu der für Vokalmusik üblichen Notenaufzeichnung in separaten Einzelstimmen, die es sehr mühsam machte, die Gesamtkomposition zu lesen. In der Tabulaturnotation kamen neben oder anstelle von Notensymbolen auch Buchstaben und Zahlen zum Einsatz, um eine möglichst kompakte Darstellung für polyphone Musik zu erreichen.[8]

Ab dem 14. Jahrhundert entstanden mehr und mehr Tabulaturschriften für alle zu dieser Zeit gebräuchlichen Instrumente. Es fehlte allerdings an gemeinsamen Regularien, weshalb sich nicht nur Tabulaturen verschiedener Instrumentengattungen, sondern auch Tabulaturen für das gleiche Instrument je nach Entstehungsort und -zeit stark unterscheiden konnten. Diese Unterschiede sind beispielsweise bei Orgeltabulaturen deutlich sichtbar. So verwenden englische und italienische Orgeltabulaturen Notensymbole auf Linien, spanische Orgeltabulaturen setzen stattdessen Ziffern ein, die bestimmten Tonhöhen zugeordnet und auf Linien für die einzelnen Stimmen angeordnet sind, während die deutsche Schreibweise Tonbuchstaben nutzt. Bei letzterer wird wiederum unterschieden zwischen einer „alten“ Form, die die Tonbuchstaben mit in Mensuralnotation ausnotierten Stimmen kombiniert, und einer „neuen“ Form, die nur Tonbuchstaben verwendet und ganz ohne Notenlinien auskommt.[9] ABBILDUNG 2 zeigt ein Beispiel für jede dieser Notenschriften.

7 Johannes Wolf, *Handbuch der Notationskunde. II. Teil: Tonschriften der Neuzeit. Tabulaturen, Partitur, Generalbaß und Reformversuche* (= Kleine Handbücher der Musikgeschichte nach Gattungen 8), Leipzig 1919.
8 Wolf, *Handbuch* (wie Anm. 7).
9 Ebd.

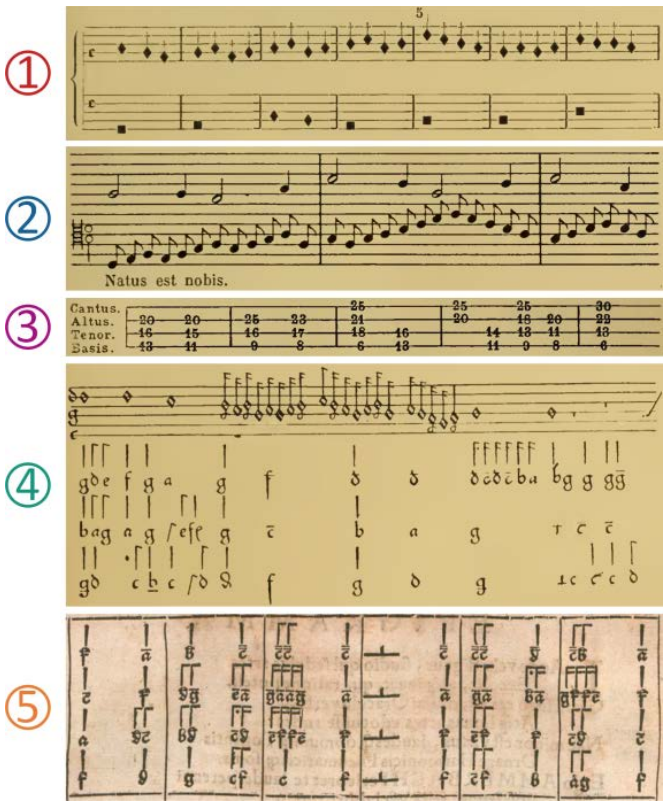

ABBILDUNG 2: GEGENÜBERSTELLUNG VERSCHIEDENER ORGELTABULATURNOTATIONSWEISEN: (1) ITALIENISCH, (2) ENGLISCH, (3) SPANISCH, (4) DEUTSCH (ALTE SCHREIBWEISE), (5) DEUTSCH (NEUE SCHREIBWEISE). ((1-4) *Handbuch der Notationskunde II. Teil,*[10] (5) *Orgel oder Instrument Tabulaturbuch*[11])

2.2. Neue Deutsche Orgeltabulatur

Die Neue Deutsche Orgeltabulaturschrift, die als bekanntester Vertreter von Tabulaturschriften für Tasteninstrumente gilt, entstand im späten 16. Jahrhundert und fand bis ins 18. Jahrhundert Anwendung.[12]

Ihre Platz und Papier sparende Buchstabennotationsweise war deutlich flexibler einsetzbar als andere Arten von Tabulaturen, da sie keine auf ein spezielles Instrument ausgelegten Schreibweisen, wie Griffbilder bei Tabulaturen für Saiteninstrumente, verwendete. Daher fand sie zunehmend auch über Orgelmusik hinaus Anwendung. Insbesondere im 17. Jahrhundert wurden viele Instrumental- oder Vokalkompositionen in die Orgeltabulaturschrift übertragen, ein Prozess, der als Intavolierung oder Intabulierung bezeichnet wird. Ein Beispiel dafür sind die geistlichen Vokalkonzerte Buxtehudes, die fast ausschließlich als Orgeltabulaturen überliefert sind.[13]

Die Neue Deutsche Orgeltabulatur wurde aber nicht nur zur Verbreitung freier populärer Kompositionen eingesetzt, sondern auch in der Ausbildung von Organisten im 17. und 18. Jahrhundert. Das prominenteste Beispiel ist Johann Sebastian Bach, der

10 Wolf, *Handbuch* (wie Anm. 7).
11 Ammerbach, *Orgel oder Instrument Tabulaturbuch* (wie Anm. 6).
12 Wolf, *Handbuch* (wie Anm. 7) sowie Cecil Warren Becker, *A Transcription of Elias Nikolaus Ammerbach's Orgel oder Instrument Tabulaturbuch*, Diss. University of Rochester 1963.
13 Hugo Riemann, *Studien zur Geschichte der Notenschrift*, Leipzig 1878.

während seines Unterrichts bei Georg Böhm die Orgeltabulaturnotation für Transkriptionen von Werken Dietrich Buxtehudes und Johann Pachelbels verwendete.[14] Ein weiteres Beispiel für die Verwendung von Orgeltabulaturen ist der in ABBILDUNG 3 abgebildete Bach'sche Orgelsatz zum Choral *Der Tag, der ist so freudenreich* (BWV 605), der eigentlich in heutiger Liniennotation festgehalten ist, bei dem Bach die letzte Zeile jedoch in Orgeltabulaturschrift niederschrieb, vermutlich da sie nur in dieser platzsparenden Form noch auf die Seite passte.[15]

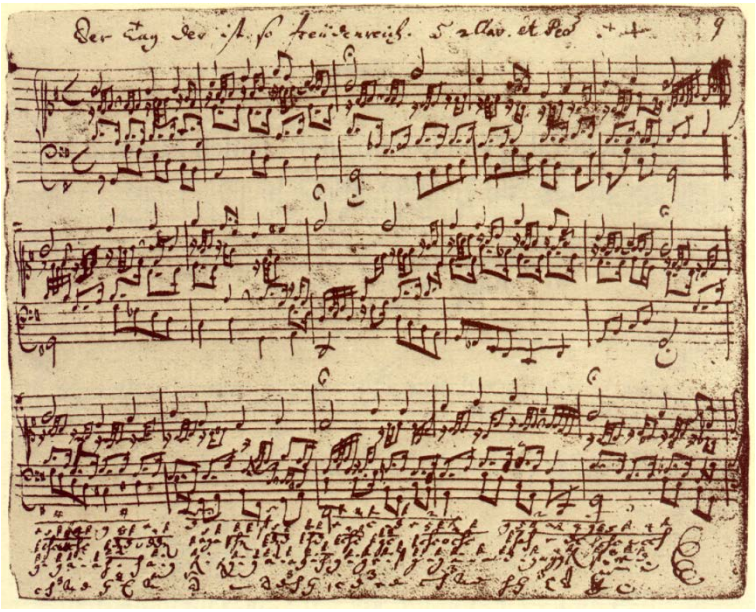

ABBILDUNG 3: NOTENBEISPIEL *DER TAG DER IST SO FREUDENREICH* VON JOHANN SEBASTIAN BACH, BWV 605 AUS DEM JAHR 1713. Die letzte Zeile ist im Gegensatz zum Rest des Stücks als Orgeltabulatur notiert (File: BWV605.png, Wikimedia Commons[16]).

Im deutschsprachigen Raum sind einige Sammlungen vorrangig handschriftlicher Orgeltabulaturnoten in der Neuen Deutschen Orgeltabulaturschreibweise erhalten geblieben. Beispielhaft seien die Klagenfurter Orgeltabulatur[17] (Entstehung ca. 1560), die Rendsburger Tabulatur[18] (1724), die Neustädter[19] und die Lüneburger Orgeltabulatur[20] (beide 17. Jahrhundert) genannt. Diese Sammlungen enthalten viele Werke von Komponisten, die über ihren Schaffensort hinaus fast gänzlich unbekannt blieben; einige Stücke sind auch anonym überliefert. Daneben existieren jedoch auch einige größere, meist gedruckte Tabulaturbücher einzelner Urheber, wie Dietrich Buxtehude, Jakob

14 Michael Maul / Peter Wollny (Hg.), *Weimarer Orgeltabulatur: die frühesten Notenhandschriften Johann Sebastian Bachs sowie Abschriften seines Schülers Johann Martin Schubart* (= Documenta musicologica II/39), Kassel, New York 2007 .

15 Becker, *Transcription* (wie Anm. 12).

16 Wikimedia Commons, File:BWV605.png. https://commons.wikimedia.org/w/index.php?title=File:BWV605.png&oldid=273414832 [letzter Zugriff am 16.02.2022].

17 Manfred Novak: „Die Klagenfurter Orgeltabulatur", in: *Wissenschaftliches Jahrbuch der Tiroler Landesmuseen* 5 (2012), S. 79–89. https://www.zobodat.at/pdf/WissJbTirolerLM_5_0079-0089.pdf [letzter Zugriff am 16.02.2022].

18 Peter Gerritz, *Choralvorspiele aus der Rendsburger Orgeltabulatur* (= Musik zwischen Nord- und Ostsee 1), Hamburg 2013.

19 Harald Wießner, *Neustädter Orgeltabulatur*, 2016. http://sphairosaudio.de/neustaedter-orgeltabulatur/ [letzter Zugriff am 16.02.2022].

20 Ebd.

Paix, Johannes Rühlig, Bernhard Schmid oder dem bereits erwähnten Elias Nicolaus Ammerbach.

Aufgrund ihrer unterschiedlichen Notation und der schwindenden Kenntnis von Orgeltabulaturen wurden einige Tabulaturen in Archiven zunächst nicht einmal als Musikstücke erkannt. So wurde beispielsweise die älteste Handschrift Johann Sebastian Bachs, die Weimarer Orgeltabulatur,[21] lange Zeit als kabbalistisches Werk angesehen und daher dem Bereich der Theologie zugeordnet.

In seinem Handbuch der Notationskunde[22] gibt Johannes Wolf auf den Seiten 32 bis 35 einen chronologischen Überblick über die wichtigsten Orgeltabulaturwerke. Einige dieser Tabulaturen sind inzwischen auch in digitaler Form über Webseiten wie der des International Music Score Library Projects (IMSLP) öffentlich zugängig.[23]

2.3. Musikwissenschaftliche Bedeutung

Mit dem Niedergang der Kirchenmusik im 18. Jahrhundert verschwand auch die Orgeltabulaturnotation aus dem Blickfeld der Organisten. Erst in der zweiten Hälfte des 19. Jahrhunderts begann eine Zeit der Wiederentdeckung alter Musik, insbesondere der Vokalmusik in Mensuralnotation,[24] durch die auch das musikwissenschaftliche Interesse an der Notation der Orgeltabulatur wieder geweckt wurde. Seitdem bilden die Übertragungen alter Notenschriften in die heutige Notation auf Basis musikwissenschaftlicher Quellenforschung das Rückgrat der Spiel- und Musiziervorlagen für Organisten.[25] In den Ausbildungen sind Orgeltabulaturen hingegen nur in wenigen Ausnahmefällen zu finden.

Am 6. Dezember 2017 hat die UNESCO den Orgelbau und die Orgelmusik in Deutschland in die Liste des immateriellen Kulturerbes der Menschheit aufgenommen.[26] Damit gewinnen auch die Identifizierung und philologisch korrekte Übertragung von Orgeltabulaturen noch stärker an Bedeutung.

Für die wissenschaftlich eindeutig identifizierten Tabulaturquellen existieren bislang jedoch nur teilweise Transkriptionen in die heutige Notation. Von den erhaltenen Intabulationen von Vokalmusik ist beispielsweise ein erheblicher Teil bislang gar nicht oder nur teilweise transkribiert worden.[27] Es gibt auch Werke, die bisher unbekannte Sammlungen erschließen, dies aber ohne eine vollständige quellenkritische Übertragung tun.[28] Darüber hinaus sind die Probleme der Transkription und Rekonstruktion von in Neuer Deutscher Orgeltabulaturschrift erhaltenen Handschriften nur in Einzelfällen, wie in Warschau[29] oder Prag,[30] näher untersucht worden.

21 Maul / Wollny, *Weimarer Orgeltabulatur* (wie Anm. 14).
22 Wolf, *Handbuch* (wie Anm. 7).
23 https://imslp.org/wiki/Category:Tablature_(keyboard).
24 Heinrich Bellermann, *Die Mensuralnoten und Taktzeichen des XV. und XVI. Jahrhunderts*, Berlin 1858.
25 Willi Apel, *Geschichte der Orgel- und Klaviermusik bis 1700*, Kassel und Basel 1967; und Willi Apel, *Die Notation der polyphonen Musik 900–1600*, Wiesbaden, Leipzig 2006[5].
26 German UNESCO Commission, *Jahrbuch der Deutschen UNESCO-Kommission 2017-2018*, 2018, S. 95.
27 Motnik, *Deutsche Tabulatur* (wie Anm. 3).
28 Wojnowska, *Thematic Catologue* (wie Anm. 3).
29 Marta Hulková, „Central European Connections of Six Manuscript Organ Tablature Books of the Reformation Era from the Region of Zips (Szepes, Spiš)“, in: *Studia Musicologica* 56.1 (März 2015), S. 3–37. DOI: https://doi.org/10.1556/6.2015.56.1.1.

2.3.1. Herausforderungen bei der Übertragung von Orgeltabulaturen

Selbst wenn Transkriptionen von Orgeltabulaturen in die heutige Notenschrift existieren, sind sie nicht immer einheitlich, transparent und philologisch korrekt. Von Ammerbachs *Orgel oder Instrument Tabulaturbuch* existieren beispielsweise Übertragungen von Cecil Warren Becker[31] und Hans-Thomas Müller-Schmidt,[32] die sich an einigen Stellen gravierend unterscheiden. ABBILDUNG 4 zeigt ein Beispiel für eine solche Abweichung. Hier oktaviert Müller-Schmidt in den Takten 2 und 3 die Alt-Stimme, während sie bei Becker in der Lage, die das Original vorgibt, steht. Dafür wird der zweite Ton der Bass-Stimme in Takt zwei bei ersterem wie im Original als „B" transkribiert, bei letzterer jedoch als „A".

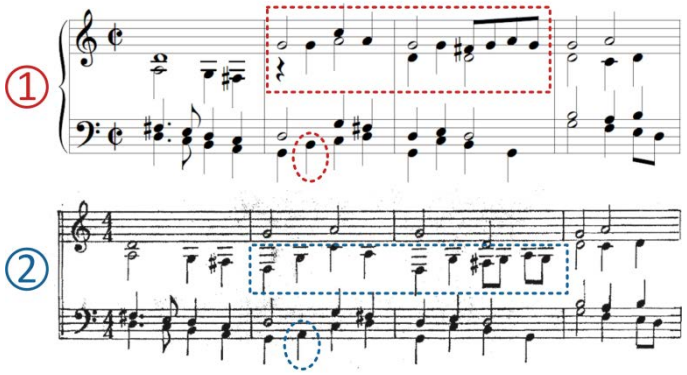

ABBILDUNG 4: EIN BEISPIEL FÜR ABWEICHUNGEN DER TRANSKRIPTION VON AMMERBACHS *ORGEL ODER INSTRUMENT TABULATURBUCH* DURCH MÜLLER-SCHMIDT[33] (1) UND BECKER[34] (2).

Unterschiede wie diese können teilweise auf Uneindeutigkeiten aufgrund von Alterserscheinungen des Originaldokuments zurückgeführt werden, in den meisten Fällen sind sie aber eine Folge des individuellen Wissens und des Transkriptionsansatzes der jeweiligen Autoren, die beispielsweise Schreibfehler in den Tabulaturen korrigieren oder Interpretationen und Änderungen zur besseren Spielbarkeit vornehmen. Dies zeigt die Wichtigkeit von einheitlichen Transkriptionsmethoden für Orgeltabulaturen, wie sie beispielsweise für die Musik des 15. und 16. Jahrhunderts[35] existieren.

30 Martin Horyna, „Medieval Organ Tablature on a Manuscript Fragment from the National Museum Library", in: *Musicalia* 10.1-2 (Dezember 2018), S. 6–42. DOI: https://publikace.nm.cz/en/periodicals/mjotcmom/10-1-2/medieval-organ-tablature-on-amanuscript-fragment-fromthe-national-museum-library [letzter Zugriff am 16.02.2022].

31 Becker, *Transcription* (wie Anm. 12).

32 Hans-Thomas Müller-Schmidt, *Orgel oder Instrumenttabulaturbuch 1583 von Elias Nikolaus Ammerbach*, 2017. https://imslp.org/wiki/Special:ReverseLookup/505757 [letzter Zugriff am 16.02.2022].

33 Ebd.

34 Becker, *Transcription* (wie Anm. 12).

35 Yu-Hui Huang et al., „Automatic Handwritten Mensural Notation Interpreter: From Manuscript to MIDI Performance", in: *Proceedings of the 16th International Society for Music Information Retrieval Conference*, Málaga 2015, S. 79–85; Jorge Calvo-Zaragoza / David Rizo / José Manuel Inesta Quereda, „Two (Note) Heads are Better Than One: Pen-Based Multimodal Interaction with Music Scores", in: *Proceedings of the 17th International Society for Music Information Retrieval Conference*, New York 2016, S. 509–514; Jorge Calvo-Zaragoza / Alejandro H. Toselli / Enrique Vidal, „Handwritten Music Recognition for Mensural Notation with Convolutional Recurrent Neural Networks", in: *Pattern Recognition Letters* 128 (2019), S. 115–121. DOI: https://doi.org/10.1016/j.patrec.2019.08.021.

2.3.2. Vorteile einer automatischen Transkription

Das von uns entwickelte *DeepTab* nimmt eine automatische Transkription von Orgeltabulaturen vor. Dadurch wird zum einen der immense Zeitaufwand einer manuellen Übertragung eingespart und zum anderen ein standardisiertes Ergebnis garantiert. Durch diese Art der Automatisierung lässt sich die Zahl der für musikwissenschaftliche Analysen verfügbaren Notationsbeispiele deutlich erhöhen. Dies stellt eine große Unterstützung für die Forschung auf diesem Gebiet dar.

Aufgrund der großen Vielfalt von Orgeltabulaturschriften in Layout und Zeichensatz ist es jedoch sehr schwierig, eine Transkriptionsmethode zu entwickeln, die auf alle Arten von Orgeltabulaturen anwendbar ist. Insbesondere handschriftliche Manuskripte stellen aufgrund des weniger strikten Layouts und der Variabilität der Ausprägungen desselben Zeichens eine sehr große Herausforderung für automatisierte Analysen dar. *DeepTab* konzentriert sich daher auf zwei Veröffentlichungen gedruckter Orgeltabulaturen: das *Orgel oder Instrument Tabulaturbuch*[36] sowie *Ein new künstlich Tabulaturbuch*,[37] beide vom deutschen Organisten und Arrangeur Elias Nikolaus Ammerbach.

Ammerbachs Tabulaturbücher gehören zu den ersten gedruckten Werken in Neuer Deutscher Orgeltabulaturschreibweise und ermöglichen aufgrund der großen Einheitlichkeit der Notation eine transparente Transkription. Darüber hinaus bietet Ammerbach in seinen Büchern einen umfassenden Grundkurs für Organisten und beschreibt in den Vorworten alle relevanten Themen, bis hin zur Griffweise. Aufgrund ihrer weiten Verbreitung im deutschsprachigen Raum sind diese Tabulaturbücher für viele Organisten prägend gewesen. Daher stellen diese Werke einen guten Startpunkt für die Entwicklung automatisierter Transkriptionsmethoden dar.

2.4. Struktur der Neuen Deutschen Orgeltabulatur

2.4.1. Zeichensatz

Die Neue Deutsche Orgeltabulaturschrift verwendet im Gegensatz zur heutigen Notenschrift keine Notenlinien, auf denen Notenzeichen zur Angabe der Tonhöhe positioniert werden. Stattdessen kommt eine Buchstabennotation zum Einsatz, bei der die Töne als Folge von Notennamen dargestellt werden. Das Tabulaturbuch von Ammerbach gibt die Oktavlage der Töne durch Groß- und Kleinschreibung der Notennamen sowie zusätzliche waagrechte Striche über den Notenhöhenzeichen an. Die Tondauer wird mithilfe über den Tonhöhenbuchstaben platzierter rhythmischer Zeichen angegeben, deren Aussehen an die Notenhälse und Balken der heutigen Notenschrift erinnert. Auch die Form und Gestaltung der Pausen- und Sonderzeichen kommt der heutigen Notationsweise sehr nah.[38]

36 Ammerbach, *Orgel oder Instrument Tabulaturbuch* (wie Anm. 6).
37 Elias Nikolaus Ammerbach, *Ein new künstlich Tabulaturbuch*, Nürnberg 1575. https://imslp.org/wiki/Special:ReverseLookup/286993 [letzter Zugriff am 16.02.2022].
38 Becker, *Transcription* (wie Anm. 12).

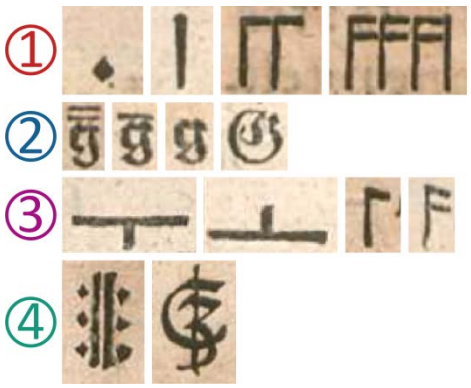

ABBILDUNG 5: EINE ÜBERSICHT ÜBER DIE VERSCHIEDENEN ARTEN VON ORGELTABULATURZEICHEN AM BEISPIEL VON AMMERBACHS TABULATURBÜCHERN: (1) NOTENDAUERZEICHEN, (2) NOTENHÖHENZEICHEN, (3) PAUSENZEICHEN, (4) SONDERZEICHEN.

ABBILDUNG 5 zeigt beispielhaft einige Tabulaturzeichen verschiedener Arten, die aus Ammerbach Tabulaturbüchern entnommen wurden (Beschriftung von links nach rechts):

1. Notendauerzeichen verschiedener Längen: Ganze Note, Halbe Note, 2 Viertelnoten, 4 Achtelnoten
2. Notenhöhenzeichen am Beispiel der Note *g* (von hoch nach tief): zweigestrichene Oktavlage, eingestrichene Oktavlage, mittlere Oktavlage (ohne Striche), tiefe Oktavlage (in Großbuchstaben)
3. Pausenzeichen verschiedener Längen: Ganze Pause, Halbe Pause, Viertelpause, Achtelpause
4. Sonderzeichen: Wiederholungszeichen, Taktwechsel (Dreivierteltakt)

2.4.2. Seitenlayout

Eine Orgeltabulaturseite besteht aus mehreren Zeilen, die durch horizontale Linien getrennt sind. Einige Herausgeber, beispielsweise auch Ammerbach, unterteilen ihre Tabulaturen zusätzlich durch vertikale Linien in Takte. Jede Zeile besteht wiederum aus mehreren Notensystemen, die mit etwas Abstand untereinander angeordnet sind, wobei jedes Notensystem genau eine Stimme der Komposition enthält.

Innerhalb eines Notensystems sind die Tabulaturzeichen in zwei Zeilen angeordnet, wie in ABBILDUNG 6 gezeigt. Die obere enthält die Notendauerzeichen *(D)*, während die untere die Notenhöhen- *(H)* sowie Pausenzeichen *(P)* enthält. Die Position der Sonderzeichen *(S)* kann je nach Herausgeber variieren, sie befinden sich jedoch üblicherweise in der oberen Zeile. Wir kürzen daher die obere Zeile im Folgenden als *D/S*-Ebene und die untere Zeile als *H/P*-Ebene ab.

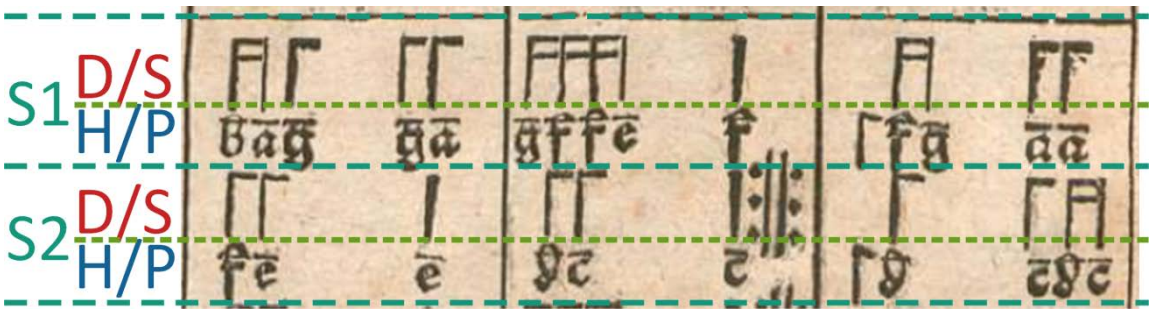

ABBILDUNG 6: DAS LAYOUT GEDRUCKTER ORGELTABULATUREN AM BEISPIEL VON AMMERBACH.[39] Jedes System *(S1, S2)* innerhalb einer Tabulaturzeile besteht aus zwei Ebenen (*Dauer-/Sonderzeichen (D/S)* und *Höhen-/Pausenzeichen (H/P)*), auf denen die Tabulaturzeichen angeordnet sind.

3. Künstliche neuronale Netze

3.1. Funktionsweise künstlicher neuronaler Netze

Deep Learning ist ein Forschungsgebiet innerhalb der künstlichen Intelligenz (KI), genauer im Bereich des maschinellen Lernens, in dem vorrangig sogenannte tiefe künstliche neuronale Netze[40] zum Einsatz kommen.

Ein künstliches neuronales Netz erlernt eine Abbildung von einer Eingabe (z. B. ein Bild oder eine Audio-Datei) zu einer Ausgabe (z. B. eine zur Eingabe passende Klassenbezeichnung). Diese Abbildung geschieht anhand von Merkmalen (*Features*), die das neuronale Netz in der Eingabe erkennt. Beispielsweise könnte für die Erkennung eines Autos in einem Bild das Erkennen von Rädern relevant sein. Dies ließe sich wiederum auf die Erkennung einer geometrischen Form mit einer bestimmten Farbe herunterbrechen.[41]

Neuronale Netze erlernen diese Merkmale implizit anhand der Eingabedaten, was oft zu deutlich besseren Ergebnissen führt als eine manuelle Vorgabe von Merkmalen. Die Erkennung von Merkmalen erfolgt hierarchisch, indem zunächst sehr einfache Merkmale (beispielsweise Kanten) gesucht und diese nach und nach zu komplexeren Repräsentationen (wie Konturen oder Objekten) kombiniert werden. Dies geschieht bei künstlichen neuronalen Netzen, indem die Eingabe eine aus mehreren Schichten bestehende Architektur durchläuft: Eine Eingabeschicht, die die Eingabe beispielsweise in Form von Pixelwerten eines Bildes darstellt, gefolgt von mehreren verborgenen Schichten, die zunehmend abstraktere Merkmale erkennen, gefolgt von einer Ausgabeschicht, über die die möglichen Ausgaben, wie etwa Objektklassen, repräsentiert werden.[42] Der Begriff *Deep Learning* leitet sich aus einer potentiell hohen Anzahl von verborgenen Schichten zwischen der Eingabe- und Ausgabeschicht ab.

39 Ammerbach, *Orgel oder Instrument Tabulaturbuch* (wie Anm. 6).
40 Michael A. Nielsen, *Neural Networks and Deep Learning*, 2015. http://neuralnetworksanddeeplearning.com/ [letzter Zugriff am 16.02.2022].
41 Ian Goodfellow / Yoshua Bengio / Aaron Courville, *Deep Learning*, 2016. http://www.deeplearningbook.org/ [letzter Zugriff am 16.02.2022].
42 Nielsen, *Neural Networks* (wie Anm. 40) und Goodfellow / Bengio / Courville, *Deep Learning* (wie Anm. 41).

Die Entwicklung von neuronalen Netzen hat sich in den letzten Jahren aufgrund von technischen Weiterentwicklungen stark beschleunigt, wodurch komplexere und auf bestimmte Problemstellungen spezialisierte Netzarchitekturen entstanden, wie beispielsweise *Convolutional Neural Networks* (CNNs) für die Verarbeitung von Bildern oder *Recurrent Neural Networks* (RNNs) zur Verarbeitung von sequentiellen Daten. Auf die Details dieser Architekturen wird an dieser Stelle nicht weiter eingegangen. Heute werden künstliche neuronale Netze in der Verarbeitung von visuellen, auditiven oder textuellen Daten eingesetzt und liefern in allen diesen Bereichen sehr gute Ergebnisse.[43]

3.2. OCR und OMR

Das Forschungsgebiet der *Optical Character Recognition* (OCR) befasst sich mit der maschinellen Erkennung von handgeschriebenem oder gedrucktem Text auf Bildern. Einsatzgebiete dafür sind unter anderem Banken (maschinelles Einlesen von Überweisungsträgern), das Gesundheitswesen (Digitalisierung von Fragebögen) oder Bibliotheken (Indizierung von Dokumenten und Bereitstellung von Suchfunktionen). OCR-Anwendungen sorgen in allen diesen Bereichen dafür, dass zeitaufwändige und fehleranfällige Prozesse automatisiert werden.[44]

Ein verwandtes Gebiet ist die *Optical Music Recognition* (OMR), welche sich mit dem maschinellen Einlesen von Musiknotationen in Dokumenten beschäftigt. Auch hier geht es darum, die komplexe und zeitaufwändige Aufgabe des manuellen Transkribierens von Noten zu automatisieren und beispielsweise alte Noten auf Papier in digitale Notenformate wie MusicXML oder LilyPond zu übertragen. Dies ermöglicht es, Musikstücke einem größeren Personenkreis zugänglich zu machen und erleichtert die Durchführung musikwissenschaftlicher Analysen, wie beispielsweise die automatisierte Suche wiederkehrender Sequenzen oder den Vergleich mehrerer Ausgaben eines Stückes.[45]
OCR- und OMR-Anwendungen verarbeiten ein Eingabebild üblicherweise in einem mehrstufigen Prozess, der sich in die im Folgenden beschriebenen Phasen einteilen lässt.[46]

---

43 Jürgen Schmidhuber, „Deep Learning in neural networks: An overview“, in: *Neural Networks* 61 (2015), S. 85–117. DOI: https://doi.org/10.1016/j.neunet.2014.09.003. arXiv: 1404.7828; sowie Nielsen, *Neural Networks* (wie Anm. 40).

44 Amarjot Singh / Ketan Bacchuwar / Akshay Bhasin, „A Survey of OCR Applications“, in: *International Journal of Machine Learning and Computing* 2.3 (2012), S. 314–318. DOI: https://doi.org/10.7763/ijmlc.2012.v2.137; Sarika Pansare / Dhanshree Joshi, „A Survey on Optical Character Recognition Techniques“, in: *International Journal of Science and Research (IJSR)* 3.12 (2012), S. 1247–1249; Monica Patel / Shital P. Thakkar, „Handwritten Character Recognition in English: A Survey“, in: *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)* 4.2 (2015), S. 345–350. DOI: https://doi.org/10.17148/ijarcce.2015.4278.

45 Jorge Calvo-Zaragoza / Gabriel Vigliensoni / Ichiro Fujinaga, „Document Analysis for Music Scores via Machine Learning“, in: *Proceedings of the 3rd International workshop on Digital Libraries for Musicology - DLfM* 2016. New York 2016, S. 37–40. DOI: https://doi.org/10.1145/2970044.2970047; Jan Hajič et al., „Towards full-pipeline handwritten OMR with musical symbol detection by U-NETS“, in: *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018* (2018), S. 225–232; Jorge Calvo-Zaragoza / Jan Hajič Jr. / Alexander Pacha, „Understanding Optical Music Recognition“, in: *ACM Computing Surveys* 53.4 (2020). DOI: https://doi.org/10.1145/3397499.

46 Pansare / Joshi, *Survey* (wie Anm. 44); Muna Ahmed Awel / Ali Imam Abidi, „Review on Optical Character Recognition“, in: *International Research Journal of Engineering and Technology (IRJET)* 06.6 (2019), S. 3666–3669; Patel / Thakkar, *Handwritten Character Recognition* (wie Anm. 44); Ana Rebelo et al., „Optical Music Recognition: State-of-the-art and Open Issues“, in: *International Journal of*

3.2.1. Vorverarbeitung

Zunächst werden die Eingabebilder vorverarbeitet, um die relevanten Vordergrundelemente (Text, Noten) besser von irrelevantem Hintergrund unterscheiden zu können. Übliche Vorverarbeitungsschritte sind die Korrektur der Bildausrichtung und Reduzierung von Verzerrungen (engl. *Deskewing* bzw. *Dewarping)*, die Reduzierung von Bildrauschen durch Anwendung eines Glättungsfilters (engl. *Noise Removal)* und die Trennung von Vorder- und Hintergrund anhand eines Schwellwertes (engl. *Binarization*).

3.2.2. Erkennung

Nach der Vorverarbeitung wird die eigentliche Zeichenanalyse durchgeführt. Seit einigen Jahren werden dafür primär tiefe neuronale Netze, insbesondere CNNs, eingesetzt. Im konkreten Aufbau der neuronalen Netzarchitekturen gibt es jedoch zwischen verschiedenen Modellen größere Unterschiede.

Man kann insbesondere zwei grundlegende Ansätze zur automatischen Zeichenerkennung unterscheiden:

Segmentierung und Klassifikation. Beim ersten Ansatz werden die Eingabebilder durch ein neuronales Netz zunächst in einzelne Objekte (z. B. einzelne Buchstaben oder Musiksymbole) segmentiert, die anschließend von einem zweiten neuronalen Netz klassifiziert werden. In diesem Fall werden für das Training Daten benötigt, in denen nicht nur die abgebildeten Zeichen, sondern auch ihre genauen Positionen im Bild durch sogenannte Bounding Boxes angegeben sind. Die Analyse einer Eingabe erfolgt in diesem Fall unabhängig Zeichen für Zeichen, und die Ergebnisse werden erst im Nachgang zu einer einzigen Ausgabe kombiniert. Der Nachteil dieses Ansatzes ist jedoch, dass durch die Einzelanalyse viele semantische Informationen verloren gehen, die insbesondere bei der Notenerkennung sehr wichtig sind. Die Tonhöhe beispielsweise lässt sich nicht aus einem Symbol allein ableiten, sondern wird durch die Position des Notenkopfes im Notensystem bestimmt. Diese Zusammenhänge müssen für die Zusammenführung der Ergebnisse rekonstruiert werden, was zusätzlichen Aufwand bedeutet. Beispiele für diesen Ansatz finden sich in den Arbeiten von Feng et al.[47] für handschriftlichen Text und Tuggener et al.[48] für gescannte Notenblätter.

Sequenz-zu-Sequenz. Ein alternatives Vorgehen sind sogenannte Sequenz-zu-Sequenz Ansätze, bei denen größere Einheiten (z. B. ganze Zeilen) auf einmal erkannt werden. Der Vorteil dieses Ansatzes ist, dass die semantischen Beziehungen zwischen den Zeichen einer gemeinsam analysierten Sequenz erhalten bleiben, was die Kom-

Multimedia Information Retrieval (IJMIR) 1.3 (2012), S. 173–190. DOI: https://doi.org/10.1007/s13735-012-0004-6; Calvo-Zaragoza / Vigliensoni / Fujinaga, *Document Analysis* (wie Anm. 45).

47 Ziyong Feng et al., „Robust Shared Feature Learning for Script and Handwritten / Machine-Printed Identification“, in: *Pattern Recognition Letters* 100 (2017), S. 6–13. DOI: https://doi.org/10.1016/j.patrec.2017.09.016.

48 Lukas Tuggener et al., „Deep Watershed Detector for Music Object Recognition“, in: *Proceedings of the 19th International Society for Music Information Retrieval Conference* (2018), S. 271–278.

bination der Ergebnisse erheblich vereinfacht. Darüber hinaus ist die Annotation der für das Training des neuronalen Netzes erforderlichen Daten weniger aufwändig, da in der Regel keine Bounding Boxen für einzelne Zeichen erforderlich sind, sondern nur die Gesamt-Zeichensequenz benötigt wird. Beispiele für diesen Ansatz sind bei Su und Lu[49] oder Dutta et al.[50] für handgeschriebene Texte und Calvo-Zaragoza, Valero-Mas und Pertusa,[51] Calvo-Zaragoza und Rizo[52] oder Alfaro-Contreras, Calvo-Zaragoza und Iñesta[53] für Noten zu finden.

3.2.3. Nachbearbeitung

In der Regel folgt nach der Zeichenerkennung ein Nachbearbeitungsschritt. Für den Fall, dass die Erkennung in kleineren Einheiten durchgeführt wurde, werden hier die Einzelergebnisse zusammengeführt. Bei der Erkennung von Noten müssen dazu die semantischen Beziehungen der erkannten Zeichen ermittelt werden (z. B. die Position einer Note im Notensystem) und in ein Datenformat übertragen werden, in dem diese Beziehungen modelliert sind. Die Notwendigkeit dieser Rekonstruktion dem semantischen Beziehungen macht die Notenerkennung komplexer als eine reine Texterkennung, bei der einzelne Zeichen weitestgehend unabhängig voneinander erkannt und transkribiert werden können.[54]

Nach dem Zusammenführen der Ergebnisse kann die Syntax und Semantik der Analyseergebnisse anhand von Wörterbüchern oder vorgegebenen Regeln untersucht werden; dabei gefundene Analysefehler können teilweise automatisch korrigiert werden. Bei der OMR werden die Ergebnisse abschließend in das gewünschte Ausgabeformat (z. B. MusicXML oder MIDI) kodiert.[55]

3.3. Transkription von Orgeltabulaturen

Die Analyse und Übertragung von Orgeltabulaturschrift war bislang kein Forschungsthema außerhalb des Bereichs der Musikwissenschaften. Außer unserer

49 Bolan Su / Shijian Lu, „Accurate Recognition of Words in Scenes without Character Segmentation Using Recurrent Neural Network“, in: *Pattern Recognition* 63 (2017), S. 397–405. DOI: https://doi.org/10.1016/j.patcog.2016.10.016.

50 Kartik Dutta et al., „Towards Accurate Handwritten Word Recognition for Hindi and Bangla“, in: *Communications in Computer and Information Science (CCIS)* 841 (2018), S. 470–480. DOI: https://doi.org/10.1007/978-981-13-0020-2_41.

51 Jorge Calvo-Zaragoza / Jose J. Valero-Mas / Antonio Pertusa, „End-to-end Optical Music Recognition Using Neural Networks“, in: *Proceedings of the 18th International Society for Music Information Retrieval Conference* (2017), S. 472–477.

52 Jorge Calvo-Zaragoza / David Rizo, „End-to-end Neural Optical Music Recognition of Monophonic Scores“, in: *Applied Sciences* 8.4 (2018). DOI: https://doi.org/10.3390/app8040606.

53 María Alfaro-Contreras / Jorge Calvo-Zaragoza / José M. Iñesta, „Approaching End-to-End Optical Music Recognition for Homophonic Scores“, in: *9th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)* 11868 (2019), S. 147–158. DOI: https://doi.org/10.1007/978-3-030-31321-0_13.

54 Calvo-Zaragoza / Hajič Jr. / Pacha, *Understanding OMR* (wie Anm. 45); sowie David Bainbridge / Tim Bell, „The Challenge of Optical Music Recognition“, in: *Computers and the Humanities* 35.2 (2001), S. 95–121.

55 Patel / Thakkar, *Handwritten Character Recognition* (wie Anm. 44); Rebelo et al., *Optical Music Recognition* (wie Anm. 46); sowie Calvo-Zaragoza / Hajič Jr. / Pacha, *Understanding OMR* (wie Anm. 45).

Arbeit[56] existieren daher noch keine weiteren Forschungsarbeiten aus dem Bereich der Informatik zu diesem Thema.

Es besteht eine große Verwandtschaft zwischen der Erkennung von Tabulaturzeichen und den Forschungsgebieten OCR sowie OMR. In allen Fällen besteht die Aufgabe darin, Zeichen eines bestimmten Alphabets auf einem Eingabebild zu erkennen und daraus eine zusammengefasste Ausgabe unter Berücksichtigung der semantischen Beziehungen der Zeichen zu erzeugen. Bei der Erkennung von Tabulaturzeichen in der Neuen Deuschen Orgeltabulaturschrift besteht dieses Alphabet aus einer Kombination von Buchstaben, die die Tonhöhe angeben, weiteren Zeichen für die Tondauer sowie Pausen und Sonderzeichen.

Viele der Erkenntnisse und Lösungsansätze aus der OCR- und OMR-Forschung lassen sich daher auch auf Orgeltabulaturen übertragen. *DeepTab* zur automatisierten Erkennung und Transkription von Orgeltabulaturen folgt in seinem Aufbau daher stark den in diesem Kapitel beschriebenen Schritten. Da künstliche neuronale Netze und insbesondere CNNs sehr gut bei allen Arten von Problemstellungen der Bildanalyse funktionieren, verwenden auch wir für den Zeichenerkennungsschritt in *DeepTab* ein künstliches neuronales Netz.

4. *DeepTab*

DeepTab analysiert gescannte Dokumente in Orgeltabulaturschrift und liefert eine Transkription in die heutige Notenschrift. Bei mehrseitigen Dokumenten werden die einzelnen gescannten Seiten zunächst aus dem Eingabedokument extrahiert und nacheinander verarbeitet. Die Transkription erfolgt in drei aufeinander folgenden Schritten mit mehreren Teilschritten:

1. Vorverarbeitung: Entzerrung der Eingabebilder und Segmentierung in einzelne Tabulaturzeilen und -systeme
2. Erkennung: Erkennung der Tabulaturzeichen in den einzelnen Systemen
3. Nachbearbeitung: Zusammenführung der Ergebnisse zu einem Gesamtergebnis und Erzeugung der Ausgabedateien

4.1. Vorverarbeitung

In diesem ersten Schritt werden die Eingabedaten für die Zeichenerkennung vorbereitet. Dies beinhaltet das Entzerren der Bilder sowie die Segmentierung in Zeilen und einzelne Tabulatursysteme.

56 Schneider et al., *Automatic Transcription* (wie Anm. 4).

4.1.1. Entzerrung

Die Qualität von Scans alter Dokumente kann sehr stark variieren. Viele Dokumente sind von Alter und Gebrauchsspuren gezeichnet, was die Lesbarkeit einschränkt. Das Papier ist oft stark vergilbt, und an manchen Stellen ist der Druck verblasst. Aufgrund der verwendeten Holzschnittdrucktechnik sind viele Seiten schief gedruckt oder erscheinen verzerrt. Dies stellt sowohl für die Analyse, als auch besonders für die Segmentierung eine Herausforderung dar. Wenn die Tabulaturzeilen aufgrund von Verzerrungen schräg verlaufen, kann es vorkommen, dass Zeichen abgeschnitten werden, wenn die Segmentierung in einzelne Tabulatursysteme durchgeführt wird. Gleichzeitig soll aber auch vermieden werden, dass unnötige Ränder hinzugefügt werden.

Um diesen Herausforderungen zu begegnen, wird als erster Vorverarbeitungsschritt ein sogenanntes *Deskewing* jeder Eingabeseite durchgeführt. Dabei werden Verzerrungen und die Rotation des Bildes ausgeglichen. Der von uns verwendete Algorithmus führt zunächst eine Liniendetektion der horizontalen Zeilentrennlinien auf der Eingabe durch. Auf den erkannten Linien werden in regelmäßigen Abständen Punkte bestimmt, für die anschließend die Verschiebung berechnet wird, die nötig ist, um die Linien horizontal auszurichten. Anhand der dabei ermittelten Parameter wird eine Transformationsmatrix erstellt, mithilfe der die Eingabegrafik entzerrt wird.

4.1.2. Segmentierung

Da es einfacher ist, die Zeichenanalyse unabhängig voneinander auf einzelnen Tabulatursystemen durchzuführen, wird die entzerrte Tabulaturseite nun in Einzelbilder für alle Systeme aufgespalten.

Zunächst erfolgt eine Aufteilung der Tabulaturseite in einzelne Zeilen anhand der horizontalen Trennlinien, die mithilfe der Liniendetektion gefunden wurden. Anschließend werden die Zeilenbilder gleichmäßig in Bilder für einzelne Systeme aufgespalten. Die Stimmenanzahl muss dabei vom Anwender oder der Anwenderin angegeben werden, da eine Schätzung durch einen Algorithmus sehr fehleranfällig ist und eine falsche Bestimmung der Anzahl der Stimmen zu einer falschen Segmentierung und daraus folgenden Problemen bei der Zeichenerkennung führen würde. Der Segmentierungsprozess wird anhand eines Beispiels in ABBILDUNG 7 veranschaulicht.

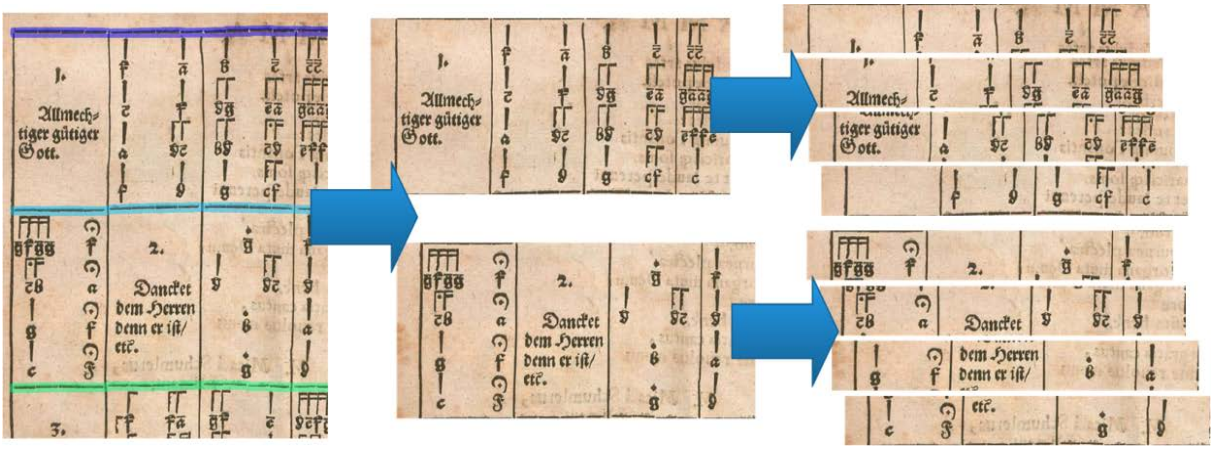

ABBILDUNG 7: SEGMENTIERUNG EINER ORGELTABULATURSEITE IN DIE ZU ANALYSIERENDEN TEILGRAFIKEN. Zunächst wird die Grafik anhand der erkannten horizontalen Linien in Einzelbilder für Tabulaturzeilen aufgespalten. Im nächsten Schritt werden diese weiter in Bilder einzelner Systeme unterteilt.

4.2. Erkennung

In diesem Schritt wird eine Tabulaturzeichenerkennung auf den entzerrten und in einzelne Tabulatursysteme segmentierten Bildern durchgeführt.

4.2.1. CSP-Netz

Der größte Unterschied zwischen der Erkennung beliebiger Textzeilen und der Erkennung von Orgeltabulaturzeichen in Bildern besteht darin, dass jedes Tabulatursystem aus zwei Zeilen besteht, auf denen die zu erkennenden Zeichen angeordnet sind. Die obere Zeile enthält die Notendauer- und Sonderzeichen (*D-/S-Ebene*), während die untere Zeile Notenhöhen- und Pausenzeichen (*H-/P-Ebene*) enthält (siehe ABBILDUNG 6).

Um diese zweizeilige Anordnung zu berücksichtigen, verwenden wir ein neuronales Zeichenerkennungsnetz mit zwei Ausgabeschichten (genannt *D/S-Output* und *H/P-Output*), die auf die Vorhersage der Zeichen der jeweiligen Ebene trainiert werden. Dies folgt der Idee des sogenannten Multitask-Lernens,[57] wobei die Aufgaben in diesem Fall die Erkennung von Zeichen zweier unterschiedlicher Zeichensätze innerhalb eines Bildes sind.

Eine solche Aufteilung der Netzwerkausgaben hat den Vorteil, dass die Codierung der Zeichensequenzen separat für die beiden Ebenen erfolgen kann. Bei einer gemeinsamen Codierung müssten alle möglichen Kombinationen von Notendauer- und Notenhöhenzeichen als separate Kennzeichnungen codiert werden, was die Anzahl der Kennzeichnungen immens vergrößern würde. Dies würde wiederum zu einem immensen Anstieg der für das Training des Netzes benötigten Anzahl an Trainingsbeispielen führen, damit jede Kombination häufig genug auftritt. Durch die Verwendung zweier

57 Rich Caruana, „Multitask Learning“, in: *Machine Learning* 28.1 (1997), S. 41–75.

Ausgabeschichten erfolgt die Erkennung der Zeichen beider Ebenen unabhängig voneinander, wodurch der Trainingsaufwand deutlich reduziert wird.

Da unser Analysenetz paarweise Zeichensequenzen liefert, nennen wir es *Character Sequence Pair Network* (CSP-Netz). Details zur verwendeten Netzarchitektur und zur Durchführung des Trainings werden in unserer Publikation[58] beschrieben.

4.2.2. Training und Anwendung des CSP-Netzes

Vor der Anwendung des CSP-Netzes wird dieses mithilfe beschrifteter Bilder von Orgeltabulatursystemen trainiert. Dabei wird eine Fehlerfunktion berechnet, die angibt, wie stark sich die für eine Eingabe an den beiden Ausgabeschichten vorausgesagten Zeichensequenzen jeweils von der korrekten Sequenz unterscheiden. Mithilfe dieses Fehlerwertes werden die Gewichtsvektoren des Netzes in einem iterativen Verfahren so angepasst, dass der Fehler im Verlauf des Trainings immer weiter reduziert wird.

Nach Abschluss des Trainings können mithilfe des CSP-Netzes Analysen von bisher ungesehenen Orgeltabulatursystemen durchgeführt werden. Das neuronale Netz liefert für jede *x*-Koordinate des Eingabebildes an jeder der beiden Ausgabeschichten eine Wahrscheinlichkeitsverteilung, die zu jedem möglichen Zeichen die Konfidenz angibt, dass dieses Zeichen an dieser Position aufgetreten ist. Aus diesen Wahrscheinlichkeitsverteilungen werden anschließend die Zeichensequenzen mit der höchsten Gesamtwahrscheinlichkeit bestimmt. Dies geschieht unabhängig voneinander für die *D/S-Ebene* und die *H/P-Ebene*. Der Erkennungsschritt resultiert also in zwei Zeichensequenzen, die anschließend zu einem kombinierten Ergebnis zusammengeführt werden müssen.

4.3. Nachbearbeitung

In diesem Schritt werden alle Ausgaben des CSP-Netzes zu einem Gesamtergebnis zusammengeführt und daraus eine LilyPond-Datei erzeugt. Aus dieser Datei wird schließlich eine grafische Ausgabe in heutiger Notenschrift generiert.

4.3.1. Ergebniszusammenführung

Das Analysenetzwerk liefert für jedes zu analysierende Tabulatursystem zwei Zeichensequenzen, eine mit Notendauer- und Sonderzeichen und eine mit Notenhöhen- sowie Pausenzeichen. Diese werden nun zu einer gemeinsamen Sequenz zusammengeführt, indem immer ein Notendauerzeichen mit einem Notenhöhenzeichen kombiniert wird, um eine Note in der heutigen Notenschrift darzustellen.

Pausenzeichen und Sonderzeichen können bei diesem Vorgang direkt in die Ergebniszeichenfolge übernommen werden. Falls aufgrund von Erkennungsfehlern keine vollständige Kombination aller Zeichen möglich ist, werden die verbleibenden Zeichen

58 Schneider et al., *Automatic Transcription* (wie Anm. 4).

einzeln zum Ergebnis hinzugefügt, jedoch mit einem *x* als Notenkopf (wenn kein Notenhöhenzeichen gefunden wurde) oder ohne Notenhals (wenn kein Notendauerzeichen gefunden wurde) formatiert, um anzuzeigen, dass in diesem Takt ein Fehler vorliegt.

Bei der Analyse mehrerer zusammengehöriger Tabulatur-Notensysteme werden nun immer die Zeichenfolgen der Notensysteme, die der gleichen Stimme zugeordnet sind, zu einer einzigen langen Folge verkettet. Die Analyse eines vierstimmigen Orgeltabulaturstücks resultiert also in vier Zeichenfolgen, unabhängig von der Anzahl der analysierten Seiten.

4.3.2. LilyPond-Ausgabe

Aus den zusammengeführten Stimmen wird nun eine LilyPond-Datei erzeugt. Beim LilyPond-Dateiformat handelt es sich um ein LaTex-artig strukturiertes Format, in dem jede Stimme eines mehrstimmigen Satzes in einem separaten Block als Zeichenfolge festgehalten wird. Dabei geben Buchstaben die Notenhöhe an (mit Kommata und Hochkommata für die Oktavlage) und Zahlen die Notenlänge. Aufgrund seiner einfachen, aber klaren Struktur eignet sich dieses Dateiformat gut für musikwissenschaftliche Analysen, insbesondere für statistische Untersuchungen.

Die Beschriftung der Trainingsdaten für das Analysenetzwerk wurde weitgehend an das Notationsschema von LilyPond angepasst, um eine möglichst direkte Übertragung der Ergebnisse nach LilyPond zu ermöglichen. So entsprechen die Notendauer-, Notenhöhen- und Pausenzeichen bis auf wenige Ausnahmen genau den in LilyPond verwendeten Bezeichnungen (z. B. 4 für eine Viertelnote, d” für ein zweigestrichenes *d* oder r8 für eine Achtelpause). Lediglich für Sonderzeichen und Fermaten, die einen komplexeren Befehl in LilyPond erfordern, kommen abkürzende Schreibweisen zum Einsatz, die während des Zusammenführungsschrittes durch die entsprechenden LilyPond-Befehle ersetzt werden. So erhalten wir für jede Stimme eine Zeichenfolge im LilyPond-Format.

Um eine vollständige LilyPond-Datei aus den ermittelten Ergebnissen zu generieren, wird eine Template-Datei verwendet, die bereits alle benötigen Layout-LilyPondbefehle enthält. In diese werden an den passenden Stellen die Zeichenfolgen für die einzelnen Stimmen eingesetzt. An anderer Stelle werden die Stimmen dann entsprechenden Notensystemen zugewiesen, so dass sich beispielsweise ein vierstimmiger Satz ergibt. Aus dieser LilyPond-Datei wird abschließend eine grafische oder digitale Notenausgabe im gewünschten Format (pdf, png, svg, midi) erzeugt.

DeepTab ermöglicht es also, zu einer pdf-Datei mit eingescannten Seiten in Orgeltabulaturschrift durch Durchlaufen der beschriebenen Schritte eine entsprechende Datei in heutiger Notenschrift zu erzeugen. Beispiele für auf diese Weise angefertigte Übertragungen finden sich in Unterabschnitt 6.3.

5. Orgeltabulatur-Datensatz

Da es bislang keine Datensätze zum Training eines neuronalen Netzes zur Erkennung von Orgeltabulaturen gab, haben wir im Rahmen unserer Arbeit[59] einen solchen Datensatz erstellt.

Als Quelle für diesen Datensatz dienten zwei gedruckte Orgeltabulaturbücher von Elias Nikolaus Ammerbach, die als Scans online frei zugänglich sind:

- Das *Orgel oder Instrument Tabulaturbuch*,[60] das aus 213 Seiten in Orgeltabulaturschrift besteht
- *Ein neues künstlich Tabulaturbuch*,[61] das 170 Tabulaturseiten enthält

Beide Werke verwenden denselben Zeichensatz, unterscheiden sich jedoch leicht im Layout. Die Qualität der verfügbaren Scans ist insgesamt gut, aber es gibt auch Stellen, an denen die Leserlichkeit eingeschränkt ist.

Zur Zusammenstellung des Datensatzes wurden aus beiden Büchern je 1200 Tabulatursysteme verwendet, die manuell mit den darauf abgebildeten Zeichensequenzen annotiert wurden. Diese Menge an Tabulatursystemen ist jedoch nicht ausreichend, um das neuronale Netz bestmöglich zu trainieren. Daher werden Datenaugmentierung und ein Datengenerator zur künstlichen Erzeugung von Tabulaturzeilen eingesetzt, um die Menge der verfügbaren Daten zu erhöhen.

5.1. Datenaugmentierung

Um sicherzustellen, dass das Erkennungsnetzwerk gut generalisiert und nicht nur die während des Trainings gesehenen Bilder auswendig lernt, wird ein Verfahren namens Datenaugmentierung eingesetzt. Dabei werden zufallsbasiert optische Änderungen an den Eingabebildern vorgenommen und so mehrere Varianten jedes Bildes erzeugt.

Die verwendeten Augmentierungsoperationen umfassen zufällige Änderungen der Bildparameter Farbe, Helligkeit und Kontrast, zufällige Verzerrungen, Skalierungen, Drehungen und das Hinzufügen von Bildrauschen durch Setzen einiger Pixel auf zufällige Farbwerte. Darüber hinaus kann der gesamte Bildinhalt leicht in eine zufällige Richtung verschoben werden. ABBILDUNG 8 zeigt die eingesetzten Augmentierungsoperationen anhand eines Beispiels.

59 Schneider et al., *Automatic Transcription* (wie Anm. 4).
60 Ammerbach, *Orgel oder Instrument Tabulaturbuch* (wie Anm. 6).
61 Ammerbach, *Ein new künstlich Tabulaturbuch* (wie Anm. 37).

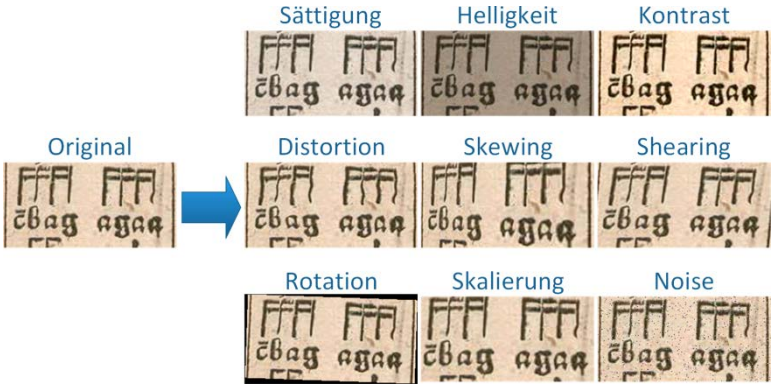

ABBILDUNG 8: VISUALISIERUNG DER VERSCHIEDENEN DURCHGEFÜHRTEN AUGMENTIERUNGSOPERATIONEN ANHAND EINES BEISPIELS. Es werden die Bildparameter Sättigung, Helligkeit und Kontrast variiert, Verzerrungen (*Distortion*, *Skewing*, *Shearing*) sowie Rotation und Skalierung durchgeführt und Bildrauschen (*Noise*) hinzugefügt.

5.2. Trainingsdatengenerator

Der von uns entwickelte Datengenerator erzeugt Bilder von Orgeltabulaturzeilen, die denen in Ammerbachs Tabulaturbüchern ähneln, indem er Bilder einzelner Tabulaturzeichen zufallsbasiert, aber nach gewissen syntaktischen Regeln, anordnet. So werden Notendauer- und Notenhöhenzeichen immer in Kombination miteinander platziert, aber die Auswahl der Zeichen selbst erfolgt rein zufällig. Es entstehen also keine logischen Melodieverläufe, es wird keine rhythmische Struktur beachtet und es werden keine harmonischen Regeln zwischen den Stimmen befolgt; kurzum es werden keine semantischen Beziehungen berücksichtigt, weder in derselben Stimme noch zwischen den Stimmen. Für das Training des neuronalen Netzes stellt dies jedoch kein Problem dar, im Gegenteil: es sorgt dafür, dass das Netz lernt, Sequenzen voneinander unabhängiger Zeichen zu erkennen, weshalb es auch keine Probleme mit in den Originalen auftretenden Schreibfehlern bekommt, die semantische Regeln verletzen.

5.2.1. Funktionsweise des Datengenerators

Als Eingabe erhält der Generator eine Sammlung von Bildern aller Tabulaturzeichen aus Ammerbachs Tabulaturbüchern. Von jedem Zeichen verwenden wir 20 bis 30 Varianten, um sicherzustellen, dass die generierten Ergebnisse eine gewisse Vielfalt aufweisen. Die Bilder der einzelnen Zeichen wurden zu diesem Zweck manuell aus den Scans der Originaldokumente extrahiert und mit einem Bildbearbeitungsprogramm freigestellt. Dieser langwierige Prozess war notwendig, um sicherzustellen, dass die generierten Zeichen sich realistisch in den Hintergrund einfügen und die Ergebnisse den Originalen möglichst ähnlich sind.

Zusätzlich zu den zu erkennenden Tabulaturzeichen wurden auch einige Beispiele für Taktstriche, Seitenränder und Textsegmente extrahiert, damit auch diese Elemente in den generierten Daten repräsentiert sind. Als Hintergründe für die generierten Bilder kommen einige Bereiche aus den Tabulaturbüchern ohne Zeichen zum Einsatz.

Der Generierungsprozess beginnt mit einem leeren Bild in gewünschter Größe, auf dem zunächst ein zufällig ausgewähltes Hintergrundbild platziert wird. Dann wird das Bild taktweise mit Tabulaturzeichen gefüllt. Jeder Takt wird von zufällig ausgewählten Taktstrichen oder Sonderzeichen auf beiden Seiten eingerahmt, und in dem dazwischenliegenden Raum werden nacheinander für jede Stimme zufällige Sequenzen von Tabulaturzeichen erzeugt. Um sicherzustellen, dass Notendauer- und Notenhöhenzeichen immer in Kombination auftreten, werden diese in Gruppen erzeugt. Die Bilder der einzelnen Zeichen werden ihrem Typ entsprechend auf der *D/S-Ebene* oder der *H/P-Ebene* der entsprechenden Stimme platziert. Dieses Prozedere wird so lange wiederholt, bis die gesamte Breite des Bildes ausgefüllt ist.

Das erzeugte Bild enthält dann mehrere Takte mit mehreren Notensystemen. Abschließend wird das Bild, wie auch bei den Realdaten, in Teilbilder für die einzelnen Notensysteme aufgeteilt.

5.2.2. Vorteile des Datengenerators

Der Generator ermöglicht es, die Auftrittswahrscheinlichkeiten der Zeichen selbst vorzugeben und damit auszugleichen, dass in den Originalbildern einige Zeichen sehr häufig vorkommen, während andere stark unterrepräsentiert sind. Durch den Einsatz der Datenaugmentierung an verschiedenen Stellen im Generierungsprozess können wir zudem Bilder erzeugen, in denen Zeichen verzerrt oder verblasst erscheinen. Dies hilft, das neuronale Netz robuster gegenüber der Herausforderung schlecht lesbarer Zeichen zu machen.

ABBILDUNG 9 zeigt einen Vergleich zwischen einer originalen Tabulaturzeile, derselben Zeile mit Datenaugmentierung und zwei generierten Tabulaturen. Während Schriftart und Grundlayout identisch sind, ist das Originalbild deutlich klarer strukturiert und hat einen höheren Kontrast. Die durch die Datenaugmentierung hervorgerufenen Änderungen am Originalbild sind eher minimal. Die künstlich erzeugten Bilder bieten hingegen eine deutlich größere Variabilität und ermöglichen somit eine bessere Generalisierung des neuronalen Netzes auf vielfältigere Daten.

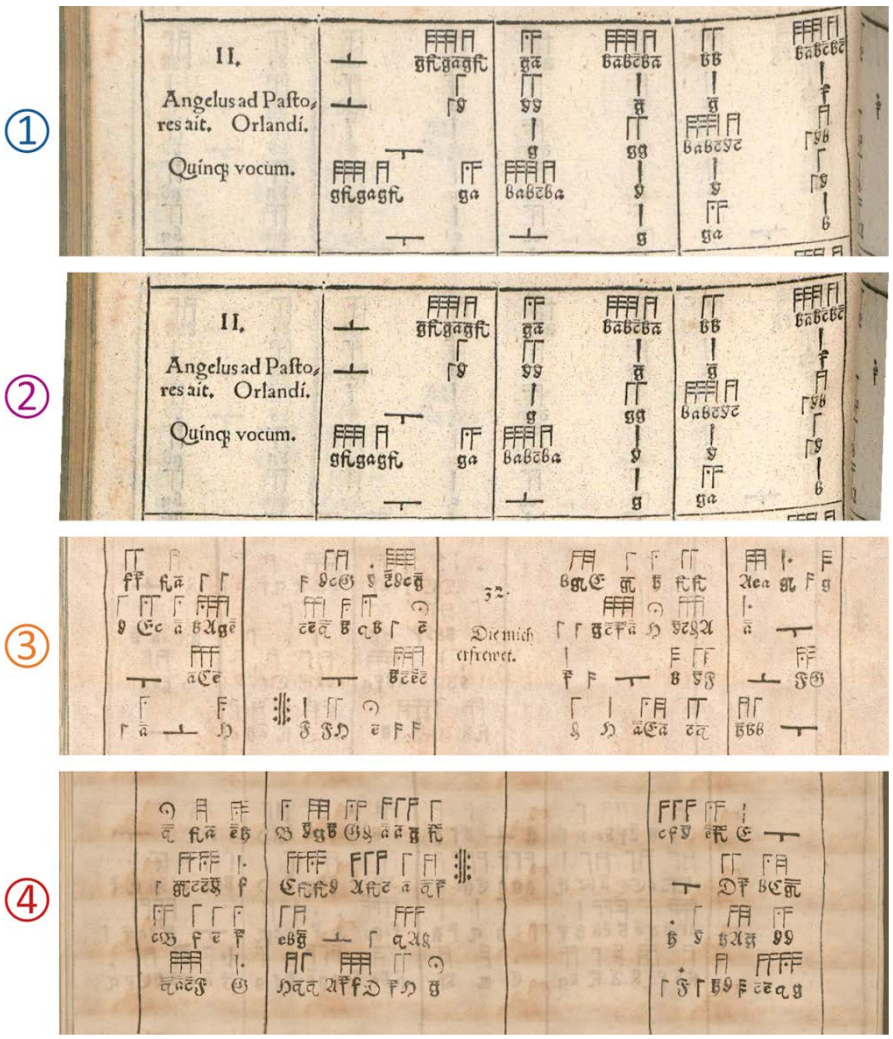

ABBILDUNG 9: VERGLEICH VON (1) EINER REALEN TABULATURZEILE AUS AMMERBACHS TABULATURBÜCHERN;[62] (2) EINER AUGMENTIERTEN VERSION DERSELBEN ZEILE; (3,4) ZWEI MITHILFE UNSERES DATENGENERATORS KÜNSTLICH GENERIERTEN TABULATURZEILEN.

5.3. Aufteilung des Datensatzes

Der aus Ammerbachs Tabulaturbüchern und zufallsgenerierten Tabulatursystemen zusammengestellte Datensatz wird in drei disjunkte Teilmengen unterteilt: Trainings-, Validierungs- und Testdaten. Die ersten beiden bestehen aus einer Mischung von echten und künstlichen Tabulaturen und kommen während des Trainingsprozesses zum Einsatz, um das neuronale Netz zu trainieren bzw. währenddessen die Performanz zu überwachen und sicherzustellen, dass es zu keiner Überanpassung auf den Datensatz kommt. Der Testdatensatz hingegen besteht nur aus Realdaten und wird in den nachfolgend beschriebenen Experimenten verwendet, um die Performanz des fertig trainierten Modells zu evaluieren. TABELLE 1 zeigt die Zusammensetzung der drei Datensatzteile aus den unterschiedlichen Quellen.

62 Ammerbach, *Ein new künstlich Tabulaturbuch* (wie Anm. 37).

Teildatensatz	*Orgel oder Instrument Tabulaturbuch*	*Ein new Künstlich Tabulaturbuch*	künstlich generiert	Summe
Training	500 (*100)	500 (*100)	20,000 (*5)	**200,000**
Validierung	200 (*25)	200 (*25)	8,000 (*5)	**50,000**
Test	500	500	0	**1,000**

TABELLE 1: DER ORGELTABULATUR-DATENSATZ, BESTEHEND AUS TRAININGS-, VALIDIERUNGS- UND TESTDATEN. Die Zahlen geben die Anzahl der Bilder an, die aus jeder Quelle stammen. Die Zahlen in Klammern geben den Faktor an, um den diese Zahl durch Datenaugmentierung vergrößert wurde.

6. Experimente

Die Qualität der mit *DeepTab* erstellten Transkriptionen wurde experimentell untersucht. Dazu wurden Metriken auf den Netzwerkvoraussagen berechnet und die in den Voraussagen auftretenden Fehler kategorisiert und diskutiert.

6.1. Evaluation des CSP-Netzwerks

Zur quantitativen Bewertung des trainierten neuronalen Netzes haben wir die folgenden Metriken auf dem aus 1000 Tabulatursystemen bestehenden Testdatensatz berechnet:

- Top-*k*-Genauigkeit: Anzahl der Bilder, bei denen die korrekte Analyse unter den Top-*k* Voraussagen war, geteilt durch die Gesamtzahl der Bilder
- Taktweise Genauigkeit: Anzahl korrekt analysierter Takte geteilt durch die Gesamtzahl der Takte
- Levenshtein-Editierdistanz: Anzahl der Änderungsschritte pro Bild geteilt durch die Gesamtzahl der Bilder
- Normalisierte Levenshtein-Editierdistanz: Anzahl der Änderungsschritte geteilt durch die Anzahl der Zeichen pro Bild, geteilt durch die Gesamtzahl der Bilder

Eine Herausforderung bei der automatisierten quantitativen Auswertung der Transkription von Orgeltabulaturen besteht darin, dass vergleichsweise lange Zeichensequenzen betrachtet werden. Da bei der Genauigkeitsmetrik nur gezählt wird, ob die vollständige Sequenz übereinstimmt, sorgt bereits ein einzelner Fehler in einer langen Zeichenfolge für eine Verringerung des Genauigkeitswertes. Daher kann die Top-1-Genauigkeit stark verzerrt sein. Wir verwenden daher zusätzlich die Top-5- und Top-10-Genauigkeit und vor allem die taktweise Genauigkeit, da einzelne Fehler hier differenzierter betrachtet werden.

Ein noch besseres Bewertungskriterium ist die Levenshtein-Editierdistanz. Sie drückt die Unterschiedlichkeit zwischen zwei Zeichenfolgen durch die Anzahl der Bearbeitungsschritte (Einfügen, Löschen, Ändern einzelner Zeichen) aus, die erforderlich sind, um die eine Zeichenfolge in die andere umzuwandeln. Um Sequenzen unterschiedlicher Länge auf verschiedenen Bildern vergleichen zu können, normieren wir die

Editierdistanz zusätzlich auf die durchschnittliche Anzahl der Änderungsschritte pro Zeichen.

Ebene	Genauigkeit				Editierdistanz	
	Top-10	Top-5	Top-1	Takt	System	Zeichen
D/S	0.996	0.996	0.970	0.993	0.055	0.00120
H/P	0.951	0.947	0.875	0.972	0.286	0.00455

TABELLE 2: EVALUATION DES CSP-NETZES AUF DEM TESTDATENSATZ.

TABELLE 2 zeigt die berechneten Metriken für Voraussagen mithilfe des CSP-Netzes auf den Bildern des Testdatensatzes. Die Metriken werden separat für die beiden Ausgaben des Netzwerks berechnet.

Das CSP-Netz erreicht insgesamt eine sehr hohe Genauigkeit. Unter den Top-5-Ausgaben des Netzes befindet sich das korrekte Ergebnis für die Notenhöhen-/Pausenzeichen in 94,7 % aller Fälle und für die Notendauer-/Sonderzeichen in 99,6 % der Fälle. Betrachtet man die taktweise Genauigkeit, so erreicht das CSP-Netz hier 97,2 % korrekt analysierte Takte für die *H/P-Ebene* und 99,3 % für die *D/S-Ebene*.

Die Levenshtein-Editierdistanz pro Notensystem beträgt 0,286 für die *H/P-Ebene*. Auf die Anzahl der Bearbeitungen pro Zeichen normiert, erhält man einen Wert von 0,00455. Das bedeutet, dass im Durchschnitt bei jedem 220. Notenhöhen- oder Pausenzeichen ein Analysefehler auftritt. Für die *D/S-Ebene* beträgt die Editierdistanz 0,055 pro Notensystem und 0,00120 pro Zeichen. Dies bedeutet, dass im Durchschnitt bei jedem 833. Notendauer- oder Sonderzeichen ein Fehler auftritt.

6.2. Fehlerbetrachtung

In diesem Abschnitt betrachten wir die Fehler genauer, die das CSP-Netz auf den Testdaten macht, und untersuchen ihre Ursachen.

Die 1000 Bilder des Testdatensatzes enthalten in Summe 105.117 Zeichen, 51.367 Notendauer- und Sonderzeichen und 53.750 Notenhöhen- und Pausenzeichen. Davon wurden von dem neuronalen Netz 258 falsch erkannt, d. h. 35 *D/S-Zeichen* und 223 *H/P-Zeichen*. Die Fehler lassen sich in die vier in TABELLE 3 dargestellten Kategorien unterteilen.

D/S		*H/P*	
Kategorie	**Anzahl**	**Kategorie**	**Anzahl**
Fehlend	21	***Fehlend***	47
Ergänzt	4	***Ergänzt***	17
Falsch	10	***Falsch***	74
		Oktaviert	85

TABELLE 3: DIE BEI DER ZEICHENERKENNUNG AUFGETRETENEN FEHLER, KATEGORISIERT IN GRUPPEN MIT DER ENTSPRECHENDEN ANZAHL AN FÄLLEN.

In vielen Fällen lassen sich die Fehler auf mangelnde Qualität des Bildmaterials zurückführen. Der Druck ist an einigen Stellen ungleichmäßig oder unvollständig, oder die Farbe ist aufgrund des Alters des Dokuments verblasst. Dadurch sind einige Zeichen undeutlich und nur schwer zu unterscheiden. An vielen dieser herausfordernden Stellen findet das neuronale Netz bereits das richtige Ergebnis, aber an anderen Stellen schlägt die Analyse fehl. ABBILDUNG 10 zeigt einige Beispiele für herausfordernde Bildausschnitte.

Die Fehlerkategorie *Fehlend* listet 68 Fälle auf, in denen Zeichen vom CSP-Netz gar nicht erkannt wurden. Die am häufigsten betroffenen Zeichen sind Taktstriche (29-mal), das Tonhöhenzeichen *e'* (5-mal) und das Zeichen für Sechzehntelnoten (3-mal). Diese Art von Fehlern tritt vor allem an Stellen auf, an denen viele Zeichen auf engem Raum erscheinen (siehe ABBILDUNG 10 Bilder 1 (falsch) und 7 (richtig)) oder Zeichen schlecht lesbar sind (siehe ABBILDUNG 10 Bilder 2 (falsch) und 8 (richtig)). Die fehlenden Taktstriche sind zum Beispiel fast immer darauf zurückzuführen, dass diese Taktstriche auf den Scans kaum sichtbar sind (siehe ABBILDUNG 10 Bild 3 (falsch)).

Die Kategorie *Ergänzt* enthält 21 Fälle, in denen Hintergrundelemente oder Textblöcke fälschlicherweise als Tabulaturzeichen erkannt wurden.

Die am häufigsten auftretende Fehlerart ist die Verwechslung von ähnlich erscheinenden Zeichen. Hier unterscheiden wir wiederum zwischen zwei Arten von Fehlern.

In die Kategorie *Falsch* fallen 84 Fälle, in denen ein völlig anderes Zeichen ausgegeben wurde als das in den Beschriftungen der Daten angegebene. Hier wurden am häufigsten die Tonhöhenzeichen für *c* und *e* (7-mal), *g* und *a* (6-mal) sowie *h* und *b* (ebenfalls 6-mal) miteinander verwechselt. Eine Verwechslung dieser Zeichen lässt sich in vielen Fällen auf einen unklaren Druck zurückführen, da sich die Tabulaturzeichen stark ähneln (siehe ABBILDUNG 10 Bilder 4, 5 (falsch) und 8, 9 (richtig)).

Die Fehler, bei denen ein korrektes Tonhöhenzeichen erkannt wurde, aber in der falschen Oktavlage, sind in der Kategorie *Oktaviert* aufgeführt. Die Oktavstriche über den Notennamen sind an einigen Stellen so schwach gedruckt, dass eine genaue Identifizierung der Oktavlage schwierig ist (siehe ABBILDUNG 10 Bilder 6 (falsch) und 9 (richtig)). Am häufigsten wurden die Tonhöhensymbole für *d* (24-mal), *c* (16-mal) und *g* (12-mal) einer falschen Oktavlage zugeordnet.

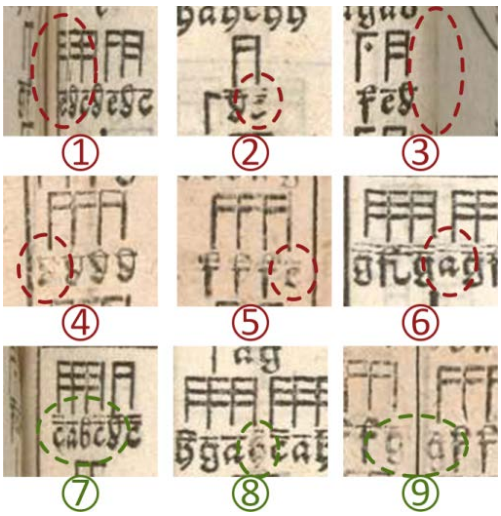

ABBILDUNG 10: BEISPIELE FÜR HERAUSFORDERNDE BEREICHE IN DEN TESTDATEN. In den Bildern 1–6 trat in den markierten Bereichen ein Analysefehler auf, während in den Bildern 7–9 selbst die schlecht lesbaren Zeichen in den eingekreisten Bereichen korrekt erkannt wurden.

Insgesamt halten sich die Fehler in Grenzen und betreffen vor allem Bildausschnitte, in denen die korrekte Zeichenerkennung mitunter auch für einen Menschen eine Herausforderung darstellt.

6.3. Beispiele für Übertragungen in die heutige Notenschrift

Die ABBILDUNGEN 11 und 12 zeigen beispielhaft die automatisch erzeugte Übertragung der ersten drei Orgeltabulaturstücke aus Ammerbachs *Orgel oder Instrument Tabulatur-buch*[63] in die heutige Notenschrift.

63 Ammerbach, *Orgel oder Instrument Tabulaturbuch* (wie Anm. 6).

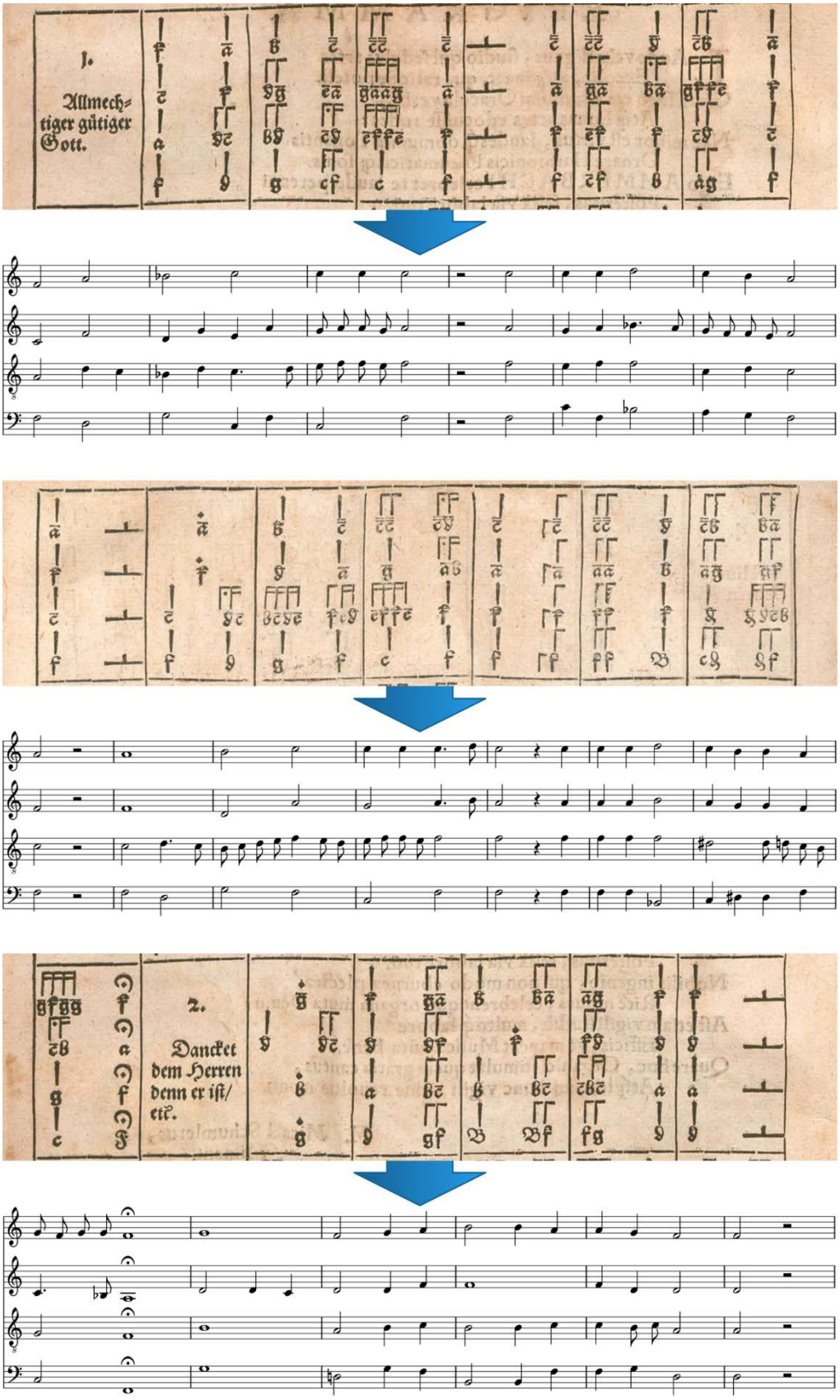

ABBILDUNG 11: DIE AUTOMATISCH ERZEUGTE ÜBERTRAGUNG DER ERSTEN DREI STÜCKE AUS AMMERBACHS *ORGEL ODER INSTRUMENT TABULATURBUCH*[64] (TEIL 1).

64 Ammerbach, *Orgel oder Instrument Tabulaturbuch* (wie Anm. 6).

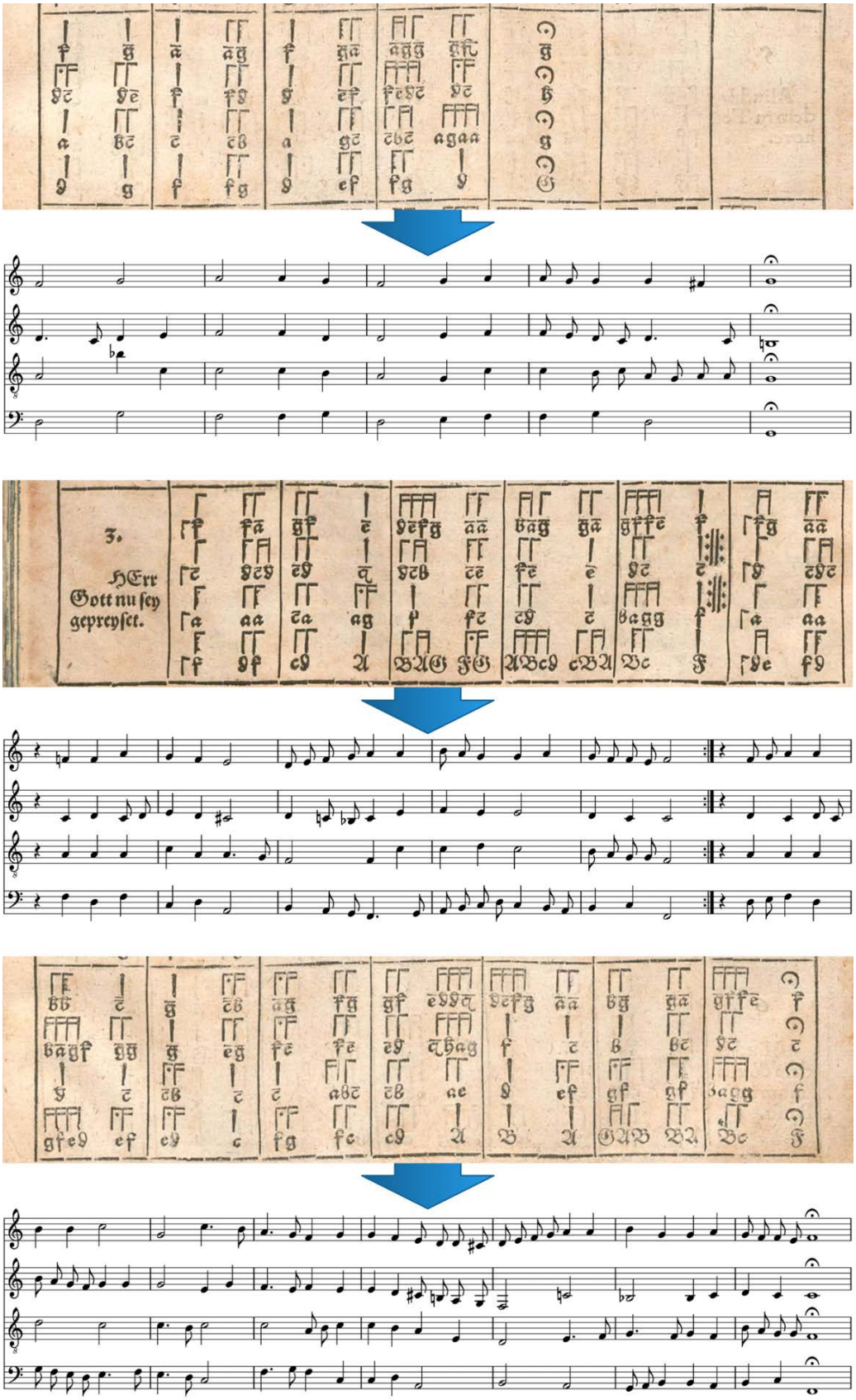

Abbildung 12: Die automatisch erzeugte Übertragung der ersten drei Stücke aus Ammerbachs *Orgel oder Instrument Tabulaturbuch*[65] (Teil 2).

65 Ammerbach, *Orgel oder Instrument Tabulaturbuch* (wie Anm. 6).

7. Zusammenfassung

In diesem Beitrag haben wir *DeepTab* präsentiert, ein auf einem tiefen neuronalen Netz basierendes Software-Werkzeug zur automatischen Transkription von Musikstücken aus der Neuen Deutschen Orgeltabulaturschrift in die heutige Notenschrift. *DeepTab* verarbeitet eingescannte Tabulaturseiten, indem es zunächst eine Segmentierung der Eingabe in einzelne Tabulatursysteme vornimmt und auf diesen anschließend mit Hilfe eines künstlichen neuronalen Netzes (genannt CSP-Netz) eine Zeichenerkennung durchführt. Das Netz liefert zu jedem Tabulatursystem zwei Ausgaben, eine für Notendauer- und Sonderzeichen und eine für Notenhöhen- und Pausenzeichen. Die daraus ermittelten Zeichensequenzen werden zu einer einzigen Ausgabe kombiniert und in das LilyPond-Notationsformat überführt. Daraus wird abschließend eine grafische Ausgabe in heutiger Notenschrift erzeugt.

Da es vor der Entwicklung von *DeepTab* keine Datensätze für das Training neuronaler Netze zur Erkennung von Orgeltabulaturen gab, wurde ein neuer Datensatz auf Grundlage zweier gedruckter Orgeltabulaturbücher von Elias Nikolaus Ammerbach erstellt. Um die Menge der Trainingsdaten zu erhöhen, wurde zudem ein Datengenerator entwickelt, der künstliche Orgeltabulaturzeilen nach dem Zufallsprinzip aus Bildern einzelner Tabulaturzeichen zusammensetzt.

Die Qualität der mithilfe des künstlichen neuronalen Netzes erzeugten Transkriptionen wurde experimentell untersucht. Das CSP-Netz erreichte bei der Analyse eines aus den Tabulaturbüchern gewonnenen Testdatensatzes eine Genauigkeit von 97,2 % und 99,3 % korrekt erkannter Takte, wobei sich der erste Wert auf die Netzausgabe für Notenhöhen- und Pausenzeichen *(H/P)* und der zweite auf Notendauer- und Sonderzeichen *(D/S)* bezieht. Im Durchschnitt tritt ein Analysefehler bei jedem 220. Notenhöhen- und Pausenzeichen und bei jedem 833. Notendauer- und Sonderzeichen auf. Viele dieser Fehler lassen sich auf die Druckqualität und das Alter der analysierten Orgeltabulaturbücher zurückführen, die an einigen Stellen eine exakte Zeichenerkennung und -differenzierung erschweren. Um ein neuronales Netz zu erhalten, das gegenüber diesen Herausforderungen noch robuster ist, wäre ein größerer Trainingsdatensatz nötig, der mehr Beispiele für solche schwierigen Fälle bietet.

Die Transkription mithilfe von *DeepTab* erfolgt ohne automatische Fehlerkorrekturen und Interpretationen, liefert also eine einheitliche Übertragung, die so nahe wie möglich am Originaldokument bleibt, was insbesondere für die musikwissenschaftliche Forschung unabdingbar ist. Bislang wurde *DeepTab* nur auf gedruckten Orgeltabulaturen trainiert und auch nur auf Tabulaturen aus zwei Tabulaturbüchern, die denselben Zeichensatz verwenden und sich nur im Seitenlayout unterscheiden. Daher eignet es sich aktuell noch nicht für eine Anwendung auf Tabulaturen, die sich stärker von den bislang verwendeten Trainingsdaten unterscheiden. Die Zusammenstellung größerer Datensätze mit verschiedenen Zeichensätzen, möglicherweise auch handschriftlichen Tabulaturen, ist daher für zukünftige Forschungsarbeiten geplant. Das langfristige Ziel ist die Entwicklung einer universell einsetzbaren Transkriptionssoftware für Neue Deusche Orgeltabulaturschrift, die Musikwissenschaftler/-innen die Erstellung philologisch korrekter Transkriptionen von Orgeltabulaturen erleichtert und dadurch eine breitere Erschließung der Orgelmusik des 16. bis 18. Jahrhunderts ermöglicht.