

Institut für Informatik
Fachbereich Mathematik und Informatik



Bachelorarbeit

Kombination maschineller Lernmethoden der Bild- und
Texterkennung auf antiken Münzdaten

Sebastian Gampe

28.03.2019

eingereicht bei:

Prof. Dott.-Ing. Roberto V. Zicari

Datenbanken und Informationssysteme (DBIS)

Erklärung zur Abschlussarbeit
gemäß § 25, Abs. 11 der Ordnung für den Bachelorstudiengang Informatik vom 06.
Dezember 2010:

Hiermit erkläre ich Herr Sebastian Gampe

Die vorliegende Arbeit habe ich selbstständig und ohne Benutzung
anderer als der angegebenen Quellen und Hilfsmittel verfasst.

Frankfurt am Main, den 28.3.2019

Sebastian Gampe

Inhaltsverzeichnis

1	Einleitung	1
	1.1 Motivation	1
	1.2 Aufgabenstellung	1
	1.3 Struktur der Arbeit	2
2	Datenbasis	3
	2.1 Online Coins of the Roman Empire (OCRE)	3
	2.2 Corpus Nummorum Thracorum (CNT)	3
3.	Grundlagen	4
	3.1 Natural Language Processing	4
	3.2 Convolutional Neural Networks	5
	3.3 Metriken der maschinellen Lernmethoden	5
	3.4 Erklärbarkeit maschineller Lernmethoden	6
	3.5 Grad-CAM	7
	3.6 Ressource Description Framework und SPARQL	9
4.	Implementierung	10
	4.1 Named Entity Recognition Pipeline	10
	4.2 Bilderkennung mittels eines Convolutional Neural Network	12
	4.3 Übertragung der Ergebnisse ins RDF-Datenmodell	14
5.	Evaluation der Ergebnisse	16
	5.1 Performance der Modelle	16
	5.2 Auswertung der NER Ergebnisse	20
	5.3 Auswertung der Bilderkennungsergebnisse	27
6.	Resümee und Ausblick	37
7.	Literaturverzeichnis	41
8.	Anhang	42
	8.1 SPARQL Abfragen	42
	8.2 Tabellen	44

1. Einleitung

1.1 Motivation

Für die Numismatik, deren Forschungsgebiet die Geldgeschichte verschiedener Zeitalter ist, sind die Münzen eine ihrer wichtigsten Quellen. Diese Münzen können zusammen mit anderen Sekundärquellen einen wichtigen Beitrag zur antiken Wirtschafts- und Kulturgeschichte leisten. Die Numismatik führt immer wieder fächerübergreifende Forschungsprojekte mit anderen Fächern durch, die den Teilnehmern neue Erkenntnisse durch das Zusammenwirken verschiedener Disziplinen ermöglicht. Eine relativ neue Entwicklung jedoch ist die Kooperation der Numismatik mit dem Fach Informatik. Da die Anwendung von Software zur Analyse von numismatischen Daten immer wichtiger wird, entstehen in diesem Zusammenhang auch neue gemeinsame Forschungsprojekte. In solch einem Rahmen ist auch die vorliegende Bachelorarbeit entstanden. Sie basiert auf der Benutzung moderner Methoden des maschinellen Lernens in der Informatik und den vorhandenen Münzdaten aus zwei numismatischen Projekten, die in der Epoche der Antike angesiedelt sind: *Online Coins of the Roman Empire* und *Corpus Nummorum Thracorum*. Der Untersuchungsgegenstand dieser Arbeit sind römische Münzen, die ein Porträt einer Person oder Gottheit und eine Inschrift auf der Vorderseite tragen. Die Identifizierung dieses Porträts wird dabei im Mittelpunkt stehen¹.

1.2 Aufgabenstellung

Die Aufgabenstellung der vorliegenden Arbeit beinhaltet eine Kombination der maschinellen Lernmethoden (ML) der Bilderkennung (*Image Recognition, IR*) und der Texterkennung (*Natural Language Processing, NLP*), welche zwei sehr unterschiedliche Herangehensweisen an das vorliegende Datenmaterial bieten. Die Bilderkennung analysiert die Eingabebilder mittels eines *Convolutional Neural Networks (CNN)*, welches in einer vorausgehenden Masterarbeit von A. Loyal darauf trainiert wurde Kaiserporträts auf römischen Münzen zu erkennen². Die Natural Language Processing Modell konzentriert sich dagegen auf die von Menschen erstellten Beschreibungen der einzelnen Münzseiten, um darin Subjekte, Objekte und Verben zu identifizieren. Dieser ML Ansatz wurde in der Bachelorarbeit von P. Klinger für diese Aufgabe trainiert³.

¹ Zu der Bedeutung römischer Porträtmünzen: Howgego 1995, 43 f.

² Loyal 2018.

³ Klinger 2018.

Die Zielsetzung dieser Arbeit gliedert sich in die folgenden Punkte:

1. Beide ML Methoden sollen auf die gemeinsame Basis der OCRE Datenbank gestellt und deren Benutzung soweit wie möglich vereinfacht werden.
2. Es soll eine Übertragung der Ergebnisse beider Ansätze von einer relationalen Datenbank in das Datenformat des *Resource Description Framework (RDF)* implementiert werden⁴. Dieser Schritt soll den folgenden Hauptteil der Arbeit ermöglichen:
3. Es soll eine detaillierte Auswertung der Ergebnisse beider Methoden erfolgen. Dies dient vor allem der Steigerung der Performance, also dem Anteil an korrekten Ergebnissen. In diesem Zusammenhang gibt es die folgende Untergliederung:
 - a) Es soll vor allem auf das Erkennen von Problemen der unterschiedlichen Ansätze auf den vorliegenden Münzdaten geachtet werden. Zusammen mit Domain-Experten sollen dann Erklärungsansätze für diese Probleme und möglicher Abweichungen in den Ergebnissen beider Methoden gefunden werden und in ein mögliches Neutraining der Modelle einfließen.
 - b) Ein weiteres Ziel soll das Erkennen möglicher Fehler in den Datensätzen der OCRE Datenbank durch den Vergleich mit den Ergebnissen von Bild- und Texterkennung mittels SPARQL Abfragen auf den erstellten RDF-Daten sein⁵. Hierbei wird grundsätzlich angenommen, dass OCRE eine sehr hohe Datenqualität besitzt und daher nur wenige Fehler gefunden werden sollten.
 - c) Da antike Münzen in unterschiedlichen Erhaltungszuständen vorliegen, sollen ebenfalls Tests an Münzen in eher schlechtem Zustand durchgeführt und deren Ergebnisse ausgewertet werden.

1.3 Struktur der Arbeit

Die vorliegende Arbeit ist folgendermaßen strukturiert: Das folgende Kapitel soll einen kurzen Überblick über die OCRE und CNT Datenbanken und ihre Stellung innerhalb des Gebiets der Numismatik bieten. Im 3. Kapitel werden die Grundlagen der verwendeten Methoden in dieser Arbeit erklärt. Kapitel 4 widmet sich dann der Implementation der vorhergehenden Arbeiten von Klinger und Loyal, um die anschließende Auswertung der neu erstellten Daten in Kapitel 5 zu ermöglichen. Eine Zusammenfassung und Ausblick auf die mögliche Fortführung der Arbeit wird im 6. Kapitel gegeben.

⁴ Resource Description Framework, <<https://www.w3.org/RDF/>> (07.03.2019).

⁵ SPARQL Protocol And RDF Query Language, <<https://www.w3.org/TR/rdf-sparql-query/>> (07.03.2019).

2. Datenbasis

2.1 Online Coins of the Roman Empire (OCRE)

Die numismatische Datenbank „*Online Coins of the Roman Empire*“ (OCRE), die in dieser Arbeit als Datenbasis dient, ist ein Gemeinschaftsprojekt der American Numismatic Society (ANS) und dem Institute for the Study of the Ancient World (ISAW) der New York University. Ziel dieses Projektes ist die Erforschung, Identifikation und Katalogisierung der von über 20 Institutionen zur Verfügung gestellten römischen Münzen. Als Grundlage hierfür dient das als Standardwerk in der Numismatik bekannte *Roman Imperial Coinage* (RIC), das die Münzen von der Gründung des römischen Kaiserreiches (31. v. Chr.) bis wenige Jahre nach dem Ende des weströmischen Reiches (491 n. Chr.) abdeckt. OCRE benutzt die vorhandene Klassifizierung der zehn Bände umfassenden RIC. Die Typenbezeichnungen enthalten den RIC Band und eine Zuweisung an Prägeherr/-ort oder porträtierte Person und eine Nummer (z.B. RIC I (second edition) Augustus 155). Die einzelnen Münzen verfügen dabei über stabile *Unified Resource Identifiers* (URI), die für das *Semantic Web* nach der *Linked Open Data* (LOD) Methodologie des Nomisma Projektes benötigt werden⁶. Die gesamte Datenbank kann online mittels eines Suchtools auf der Webseite oder dem Herunterladen eines RDF-Dumps und anschließender SPARQL Abfragen erschlossen werden⁷. In der vorliegenden Arbeit werden die ikonographischen Beschreibungen, hier Designs genannt, und die Bilder der einzelnen Münzseiten als Grundlage für NLP und Bilderkennung herangezogen. Auf den in dieser Datenbank verfügbaren Bildern der Münzvorderseiten basiert das Training des Bilderkennungsnetzes von Antje Loyal⁸.

2.2 Corpus Nummorum Thracorum (CNT)

Eine weitere Datenbank mit numismatischen Hintergrund ist die des *Corpus Nummorum Thracorum* Projektes (CNT). Im Gegensatz zu OCRE beschränkt sich CNT auf einen wesentlich kleineren geographischen Raum: das antike Thrakien, auf dem Gebiet des heutigen Bulgarien, Nordgriechenland und dem europäischen Teil der Türkei. Das Hauptforschungsgebiet ist die Veränderung des lokalen Geldsystems von der Archaik bis in die späte römische Kaiserzeit (2. Hälfte 6. Jh. v. Chr. bis 3. Jh. n. Chr.). Weiterhin sollen im

⁶ Nomisma, <<http://nomisma.org/>> (08.03.2019); Beispiel für eine solche URI mit der RIC Bezeichnung ist: <[http://numismatics.org/ocre/id/ric.1\(2\).aug.155](http://numismatics.org/ocre/id/ric.1(2).aug.155)> (08.03.2019).

⁷ Online Coins of the Roman Empire, <<http://numismatics.org/ocre/>> (08.03.2019);

⁸ Loyal 2018, 25.

aktuellen Projekt antike Münzstempel identifiziert und eine neue erweiterte Typologie für das Gebiet erstellt werden⁹. Daneben wird parallel eine Übertragung der verwendeten relationalen Datenbank in das RDF-Format entwickelt. Dieser RDF-Datensatz soll ebenfalls auf dem Nomisma Portal veröffentlicht werden. Diese Datenbank enthält die Münzdaten zahlreicher internationaler Institutionen, die in einer neuen Typologie aufgearbeitet werden. Die Münzbeschreibungen sind mit einer einfachen Nummerierung gekennzeichnet. Die Datenbank ist zurzeit extern nur durch das integrierte Suchtool zugänglich, da die RDF-Daten noch nicht veröffentlicht wurden¹⁰. P. Klinger hat die NER Pipeline auf dieser Datenbank trainiert¹¹.

3. Grundlagen

In diesem Kapitel sollen die Grundlagen der vorgelegten Arbeit erläutert werden. Zuerst wird ein Überblick über die beiden verwendeten maschinellen Lernmethoden *Natural Language Processing (NLP)* und Bilderkennung mittels *Convolutional Neural Networks (CNN)* und ihre Evaluierungsmetriken gegeben. Anschließend soll auf das Thema der Erklärbarkeit maschineller Lernmethoden und die dazu verwendete Grad-CAM Technik eingegangen werden. Zuletzt wird das Datenformat des *Resource Description Framework (RDF)* und die damit verbundene *SPARQL Protocol And RDF Query Language (SPARQL)* behandelt.

3.1 Natural Language Processing

Natural Language Processing ist eine Art der Informationsextraktion, die menschliche Sprache für den Computer verständlich machen soll. Für diesen Zweck werden maschinelle Lernmethoden wie *Feature Extraction* zum Training eines Klassifizierers, der einer Eingabe nach einer Lernphase selbstständig eine von einer endlichen Menge von Klassen zuweisen soll, verwendet. Der NLP Algorithmus zerlegt einen Text zuerst in einzelne Sätze und unterteilt diese anschließend in einzelne Wörter (*Tokenization*). Diesen einzelnen Bestandteilen des Satzes wird im nächsten Schritt ihre grammatikalische Funktion durch eine syntaktische Analyse zugewiesen (z.B. Substantiv, Verb, usw.) Der letzte Schritt besteht dann in der Identifikation von Entitäten mit Eigennamen (*Named Entities*, z.B. Personennamen, Ortsnamen, usw.). Diese Technik nennt sich wird *Named Entity Recognition (NER)* und soll

⁹ Die bisherige Typologie besteht vor allem aus einzelnen Arbeiten, die die lokalen Städteprägungen aufgearbeitet haben. Für Byzantion wäre dies beispielsweise: Schönert-Geiß, Griechisches Münzwerk. Die Münzprägung von Byzantion 1-2 (Berlin 1970–1972).

¹⁰ Corpus Nummorum Thracorum, <<https://www.corpus-nummorum.eu/about.php>> (09.03.2019).

¹¹ Klinger 2018, 3.

die Vorkommen dieser speziellen Nomina finden. Dabei wird versucht mit regulären Ausdrücken einen Satz so zu unterteilen, dass eine Entität, die aus mehreren Wörtern besteht (z.B. Konstantin der Große), nicht von ihren Bestandteilen getrennt wird (*Chunking*)¹².

3.2 Convolutional Neural Networks

Convolutional Neural Networks sind künstliche Netzwerke zur überwachten Klassifizierung von Eingabedaten. Das Vorbild hierfür sind die für die Reizübertragung im menschlichen Gehirn zuständigen Neuronen. Die einzelnen Neuronen eines künstlichen neuronalen Netzes erhalten eine Eingabe und berechnen mit Hilfe einer Eingangsfunktion und einer Aktivierungsfunktion die Ausgabe des Neurons. Ein neuronales Netzwerk ist aus mehreren miteinander verbundenen Schichten aufgebaut: Die Eingabeschicht erhält die Daten, die versteckten Schichten berechnen das wahrscheinlichste Ergebnis und die Ausgabeschicht weißt die Klasse zu. Für diese Aufgabe muss das Netzwerk aufwändig trainiert werden: Der Lernvorgang verändert die Eingangsfunktion eines Neurons. Diese berechnet aus den Eingaben, die in das Neuron fließen, eine gewichtete Summe, die als Grundlage für die Aktivierungsfunktion dient. Beim Lernvorgang wird versucht die einzelnen Gewichtungen dieser Funktion auf jedem Neuron so lange mittels einer Fehlerfunktion zu verändern, bis die Eingabedaten die korrekte Klasse als Ausgabe des Netzwerks erzeugen. Ein *Convolutional Neural Network* ist eine spezielle Art von neuronalem Netz. Diese Netzwerke können mit Hilfe einer sogenannten Faltungsoperation eine größere Performance als konventionelle neurale Netze erreichen und bieten somit einen großen Zeitvorteil im Rahmen des Trainings¹³.

3.3 Metriken der maschinellen Lernmethoden

Um die Ergebnisse eines auf maschinellen Lernmethoden basierenden Klassifizierungsverfahrens zu bewerten, werden in der Forschung üblicherweise die folgenden Metriken angewandt: *Accuracy, Precision und Recall*.

Dabei muss das trainierte Modell vorher nicht gesehene Daten klassifizieren. Dies ist bei den beiden hier als Grundlage dienenden Arbeiten durch eine Aufteilung in Trainings- und Testmenge geschehen. Da die Modelle die Daten der Testmenge nicht kennen, können hier die Metriken als Ausdruck des Vergleichs zwischen der vom Modell ermittelten Klasse und der in der Datenbank vorhandenen, sogenannten Groundtruth zum Einsatz kommen. Die

¹² Bird u. a. 2009, 261–267. 281–283; Klinger 2018, 2. 6 f.

¹³ Goodfellow – Bengio – Courville 2016, 102. 163 – 166. 321; Loyal 2018, 7–13 mit Abb. 2.4.

Klassifizierungen werden dafür in eine von vier Kategorien eingeteilt:

1. *Richtig Positiv*: Modell und Groundtruth geben Positiv aus.
2. *Richtig Negativ*: Modell und Groundtruth geben Negativ aus.
3. *Falsch Positiv*: Modell ermittelt Positiv, aber Groundtruth ist Negativ.
4. *Falsch Negativ*: Modell ermittelt Negativ, aber Groundtruth ist Positiv.

Somit beschreiben 1. und 2. die korrekt klassifizierte Datensätze, während in den Fällen 3. und 4. Fehler bei der Klassifizierung passiert sind.

Die einzelnen Metriken ermitteln jetzt aus verschiedenen Zusammensetzungen der Fälle 1. – 4. Die folgende Ergebnisse¹⁴:

Accuracy misst den Prozentsatz der korrekt klassifizierte Eingabedaten. Die Metrik *Precision* zeigt an, wie viele der gefundenen Datensätze relevant waren. *Recall* ermittelt dagegen die Prozentzahl der relevanten Datensätze, die von dem Klassifizierer gefunden wurden. Daneben gibt es noch die Metriken *F-Score* (harmonisches Mittel aus *Precision* und *Recall*) und den Top-1 und Top-5 Error, die jeweils die Fehlerrate bei der Wahrscheinlichsten oder den fünf wahrscheinlichsten Klassen zeigen¹⁵.

3.4 Erklärbarkeit maschineller Lernmethoden

Maschinelle Lernmethoden (ML) als Instrument der Bewältigung sehr großer Datenmengen werden in immer mehr Bereichen des menschlichen Lebens eingesetzt. Ein wesentliches Problem dabei ist es, die komplexen Algorithmen, Modelle und deren Ergebnisse auch Nicht-Domain Experten zu vermitteln. Selbst Menschen, die sich intensiv mit dem Thema befassen, können aufgrund der riesigen Datenmengen und den zahlreichen beim Lernvorgang ausgeführten Berechnungen kaum nachvollziehen, wie ein Modell seine Entscheidungen trifft. Auch die Komplexität der Datenbanken, die zum Lernen genutzt werden, spielt hier eine Rolle, da beispielsweise die Exemplare der zu lernenden Klassen nicht in gleicher Anzahl vorhanden sind und auch unsaubere Datensätze problematisch sind. Somit fällt es den beteiligten Wissenschaftlern schwer das Verhalten eines Modells transparent zu machen und somit Vertrauen in dessen Entscheidungen auch bei Laien zu erzeugen. Auch die Forschung selbst kann von Erklärbarkeit stark profitieren, da beim völligen Versagen eines ML Modelles nicht immer klar ist, was dieses Verhalten ausgelöst hat. Aufgrund des weiter stark zunehmenden Interesses an dieser Form der künstlichen Intelligenz und dem dagegen noch eher geringen

¹⁴ Die Formeln der Metriken können hier nachgeschlagen werden: Klinger 2018, 9 f.; Loyal 2018, 17 f.

¹⁵ Klinger 2018, 9 f.; Loyal 2018, 17 f.; Bird u. a. 2009, 239 f.; Top-1 und Top-5 Error wurden nur in Loyal 2018 genutzt.

Forschungsinteresse zur Erklärbarkeit maschineller Lernmethoden wird dieses Thema in Zukunft immer wichtiger werden. Ein Ziel muss es sein das Modell und seine Entscheidungsfindung anhand mehrerer Beispieldaten erklären zu können. Dies kann auch zu einer verbesserten Zusammenarbeit von ML Experten und den Domain Experten anderer Forschungsgebiete, in welchen diese Methoden zum Einsatz kommen sollen, führen. Weitere Ziele hierfür sind: die Verhinderung von Modellen, die ihre Entscheidungen aufgrund völlig falscher Gründe treffen; bessere Auswertbarkeit der Ergebnisse von ML basierter Forschung für Geldgeber und Verleger; Reduzierung der Kosten beim Einsatz von trainierten Modellen; bessere Wartbarkeit bei Neutraining eines Modells aufgrund neuer Daten oder Klassen; und natürlich das Erlangen neuer Erkenntnisse im Bereich der ML Forschung¹⁶.

Petkovic u. a. konnten zeigen, dass das Verständnis und somit auch das Vertrauen in die Entscheidungen eines Random-Forrest Algorithmus¹⁷ im Bereich Bioinformatik von Nicht-ML Experten bei der Verwendung eines benutzerzentrierten Ansatz zur Erklärung des Modells merklich angestiegen sind¹⁸.

3.5 Grad-CAM

Wie oben schon erwähnt wurde, gibt es ein wesentliches Problem die Modelle und die Ergebnisse von maschinellen Lernmethoden für Numismatiker als Nicht Domain Experten der künstlichen Intelligenz transparenter zu machen. Für diesen Zweck wurde eine Technik entwickelt, die es erlaubt die Ergebnisse der Bilderkennung mittels *Convolutional Neural Networks (CNNs)* besser erklären zu können: *Gradient-weighted Class Activation Mapping (Grad-CAM)*¹⁹.

Grad-CAM wurde entwickelt, um bei CNNs mit teilweise verbundenen Schichten, was dem gewählten Ansatz von Loyal auf dem diese Arbeit aufbaut entspricht, eingesetzt zu werden²⁰. Das Vorgehen dieser Methode ist es, auf jedem Bild, welches das trainierte Netzwerk durchläuft, die Pixel hervorzuheben, die zu der jeweiligen Klassifikation geführt haben. Diese Technik ist sehr stark klassendifferenziert ist, da nur die Pixel für jeweils eine Klasse

¹⁶ Petkovic u. a. 2018, 204 f.; Selvaraju u. a. 2017, 1; <<https://www.darpa.mil/program/explainable-artificial-intelligence>> (05.03.2019).

¹⁷ Random Forrest ist eine ML Lernmethode die durch mehrere Klassen-Entscheidungsbäume (den sog. Wald) eine Mehrheitsentscheidung zur Klassifizierung herbeiführt. Petkovic u. a. 2018, 205.

¹⁸ Weiterhin konnte sogar die Anzahl der verwendeten Features bei kaum absinkender Genauigkeit reduziert werden. Petkovic u. a. 2018, 212–214.

¹⁹ Selvaraju u. a. 2017, 1; Loyal 2018, 20; Die Implementierung von Grad-CAM in dieser Arbeit wurde mittels folgendem Code bewerkstelligt: GitHub, <<https://github.com/gautamMalu/caffe-gradCAM/blob/master/00-classification-gradCAM-Visualization.ipynb>> (04.03.2019).

²⁰ Loyal 2018, 11 f.

herausgestellt werden. Das Ergebnis wird dann mittels einer sogenannten *Heatmap*, die über das ursprüngliche Eingabebild (eine Münze) gelegt wird, visualisiert (siehe Abb. 1)²¹. Man kann hier deutlich erkennen, dass das Netzwerk für die Klassifikation als „Augustus“ insbesondere die Pixel im Hals- und Stirnbereich des Porträts herangezogen hat. Die Pixel des hinteren Halsbereiches sind durch die dunkle Rotfärbung als besonders wichtig angesehen worden.

Diese Heatmaps können für alle 62 trainierten Klassen des in dieser Arbeit verwendeten CNNs angelegt werden. Das bedeutet, dass man für eine einzige Münze bis zu 62 Heatmaps erhalten könnte. Aus Gründen der Praktikabilität wurde hier eine Beschränkung auf die fünf Klassen mit der höchsten Wahrscheinlichkeit, die einen *Threshold* von 5% überschreiten, festgelegt.

Um die Pixel zu bestimmen, die positiv zur Klassifizierung beigetragen haben nutzt Grad-CAM die Gradienteninformationen der letzten Faltungsschicht des Neuralen Netzwerkes: Zuerst wird das Bild dem trainierten Netz übergeben und als Ziel für die Heatmap-Erstellung die Klasse mit der höchsten Wahrscheinlichkeit festgelegt. Dabei durchläuft das Bild die versteckten Faltungsschichten des Netzwerks und am Ende werden die Wahrscheinlichkeiten für alle vorhandenen Klassen berechnet. Anschließend werden die Gradienten für fast alle Klassen auf null gesetzt. Die Klasse mit der höchsten Wahrscheinlichkeit erhält dagegen eine eins. Jetzt wird das Signal durch das soeben durchlaufene Netzwerk zu den Input-Neuronen zurückpropagiert. Im letzten Schritt fließen die gewonnenen Informationen über die wichtigsten Pixel in die zu erstellende Heatmap, die dann über das Eingabebild gelegt wird²².



Abb. 1: Heatmap einer Münze des Typs RIC I (second edition) Augustus 9A²³

Nach der Erstellung des kombinierten Bildes kann dieses den Domain-Experten vorgelegt werden, um die jetzt sichtbaren Klassifizierungskriterien bewerten zu können. Der Betrachter

²¹ Selvaraju u. a. 2017, 2.

²² Selvaraju u. a. 2017, 3 f.; Siehe dazu auch: Selvaraju u. a. 2017, 4 Abb. 2; Allerdings wird dort das Bild mit Guided Grad-Cam, einem weitergehenden Ansatz, visualisiert.

²³ Münzkabinett Berlin, 18207660 <<https://ikmk.smb.museum/object?id=18207660>> (22.03.2019).

soll hier die Bereiche der Münzen erkennen, die vom Netzwerk als besonders wichtig beurteilt wurden. Dabei stellen sich die folgenden Fragen: Nimmt das Netzwerk die Klassifizierung nur aufgrund des Porträts vor oder wird auch die Münzinschrift als Bewertungskriterium herangezogen. Welche Eigenschaften des Porträts (z.B. Frisur, Kopfschmuck, Gesicht, usw.) tragen besonders zur Klassifizierung bei? Um einen besseren Überblick über diese Kriterien zu bekommen, wird es nötig sein eine größere Reihe an Heatmaps von Münzen anzulegen. Dies wird insbesondere Auswirkungen auf die Bewertung der Arbeit des Netzes und die zukünftige Vorgehensweise im Umgang mit einem eventuellen Neutraining haben. Ein besonderes Interesse wird auch auf die Fälle gelegt werden, bei denen das Bilderkennungsmodell völlig von der vorgegebenen Groundtruth abweicht. Möglicherweise kann Grad-CAM bei diesen Exemplaren wertvolle Einsichten liefern (siehe Abschn. 5.3)²⁴.

3.6 Ressource Description Framework und SPARQL

Das *Ressource Description Framework* gehört zur Methodologie des sogenannten *Linked Open Data (LOD)* als Teil des semantischen Webs. Das RDF-Modell benutzt über *Unified Resource Identifier (URI)*, z.B. Webseiten) eindeutig identifizierbare Konzepte wie „Tetradrachme“ (Subjekt) und weist diesen mittels Eigenschaften (Prädikat) Werte (Objekt) zu. Damit entstehen Tripel, welche im RDF-Graphen mittels Knoten (Konzepte und Werte) und gerichteten, benannten Kanten (Prädikate) als Aussagen beschrieben werden²⁵. Um Konzepte und Eigenschaften zu definieren werden Ontologien wie die von Nomisma verwendet, die ein ausgefeiltes System zur Beschreibung von Münzen und damit in Verbindung stehenden Merkmalen bietet²⁶. Ein RDF-Graph kann online oder lokal mittels eines von der Onlinevariante erzeugten RDF-Dumps durch die *SPARQL Protocol And RDF Query Language* abgefragt werden. Ein sogenannter SPARQL Endpoint ermöglicht es dem Benutzer selbstständig anhand der verwendeten Ontologie den RDF-Graphen zu untersuchen²⁷. Dabei werden Variablen anstatt der Konzepte benutzt, die durch Eigenschaften mit anderen Variablen oder konkreten URI's als zu erfüllende Bedingung verbunden sind. Die Resultate der Abfrage können als Subgraph oder Tripelaufistung angezeigt werden (siehe Abb. 9)²⁸.

²⁴ Vgl. Loyal 2018, 20.

²⁵ Beispiel für eine Aussage wäre: Coin hasMint Byzantion. „Coin“ und „Byzantion“ wären demnach Knoten und „hasMint“ eine gerichtete Kante von dem Knoten „Coin“ zu dem Knoten „Mint“.

²⁶ Nomisma. Ontologie, <<http://nomisma.org/ontology>> (22.03.2019); Gruber 2018, 18–21.

²⁷ Eine Beispielabfrage ist hier zu sehen: Heath 2018, 40 mit Abb. 1. Die in dieser Arbeit verwendeten Abfragen sind in Abschnitt 8.1 aufgeführt.

²⁸ Heath 2018, 36. 38–41. 44–46.

4. Implementierung

Dieses Kapitel soll sich auf die Implementierung von NER Pipeline und Bilderkennungsmodell konzentrieren. Zu Beginn wird ein Überblick über die vorherige Implementierung der NER Pipeline gegeben und anschließend erläutert, was im Rahmen dieser Arbeit noch hinzugefügt wurde. Bei der Umsetzung des Bilderkennungsmodells wird ebenso vorgegangen. Den Schluss bildet die Vorstellung der Übersetzung der Resultate beider Ansätze in das RDF-Format.

4.1 Named Entity Recognition Pipeline

Da die vorliegende Arbeit auf der abgeschlossenen Bachelorarbeit von P. Klinger basiert, konnte ein großer Teil der vorhandenen Implementierung weiterverwendet werden. Aufgrund der hier angewandten Fragestellung wurde lediglich die auf *Natural Language Processing (NLP)* basierende *Named Entity Recognition (NER)* Pipeline, die zur Auffindung von Personen und Objekten in Münzdesigns entwickelt wurde, benutzt²⁹. Die Reihenfolge der einzelnen Schritte ist in Abb. 2 wiedergegeben.

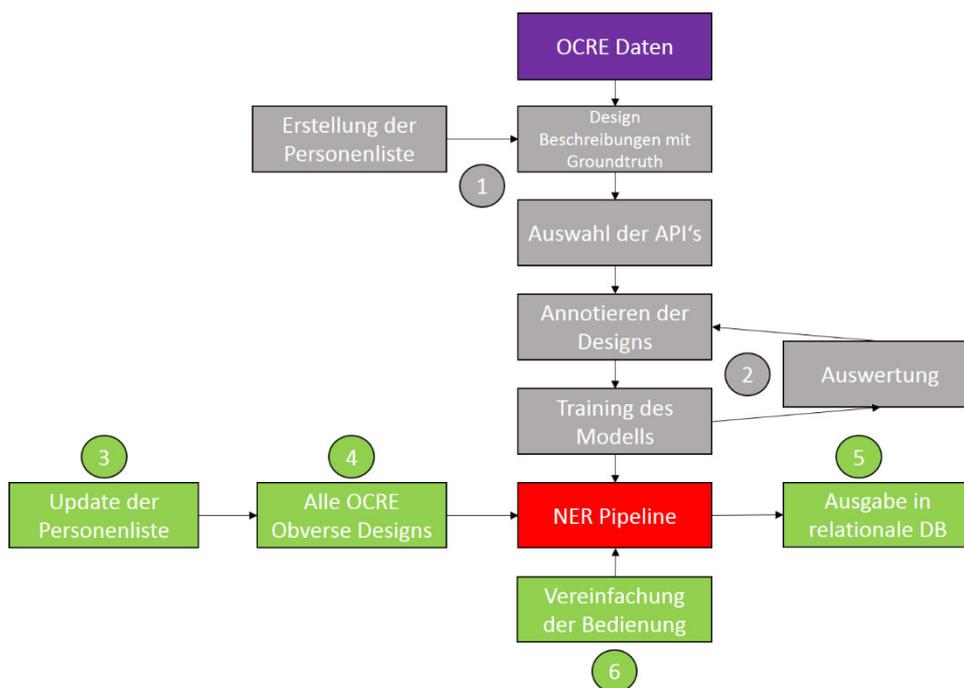


Abb.2: Ablaufdiagramm der NER Pipeline. Grau: Vorhergehende Arbeiten durch P. Klinger; Grün: Implementierungen im Rahmen dieser Arbeit.

(1) Die NER Pipeline verwendet die von den Münzbearbeitern verfassten Designs als Input und

²⁹ Klinger 2018, 12–16.

gibt Entitäten in den Kategorien PERSON, OBJECT, ANIMAL und PLANT aus (siehe Abb. 3). Hier wurden lediglich die Ergebnisse der PERSON Klassifikation verwendet.

Umgesetzt wurde die Pipeline mit dem NLP Modul *spaCy* für *Python* mit der Entwicklungsumgebung *Jupyter Notebook*³⁰. Für das Training des Modells wurde eine Liste mit Entitäten (z.B. *Augustus*, *Tiberius*, *Claudius*, usw), die in den CNT Designs vorkommen, und eine Personenhierarchie erstellt. (2) Diese Liste in einer Tabelle einer relationalen Datenbank wurde zur Annotation der Designs in den OCRE Daten verwendet. Die annotierten Designs werden als Grundlage des Trainingszyklus verwendet. Dazu werden die Daten in eine Trainings- und eine Testmenge unterteilt, damit die Performance des Modelles evaluiert werden kann (siehe Abschn. 5.1). In der Testmenge gibt es in der Trainingsmenge nicht vorkommende Entitäten, um zu prüfen, ob das Modell diese ebenfalls erkennen kann. Der letzte Schritt des Trainingszyklus besteht aus einer Auswertung der erstellten Entitätenlabel. Bei nicht der sogenannten Groundtruth entsprechenden Ergebnissen, können diese zur Optimierung der Personenliste verwendet werden (siehe Abb. 4)³¹.

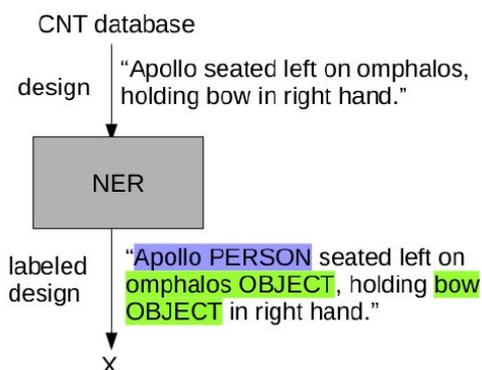


Abb. 3: Schema der NER Pipeline³²

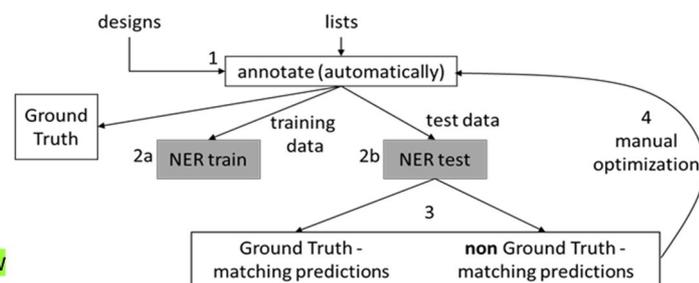


Abb. 4: Schema des NER Trainings Workflows³³

(3) Als Vorarbeit im Rahmen eines Forschungsprojektes wurde die bestehende Personenliste mit den auf den OCRE Daten neu gefundenen Entitäten und weiteren für OCRE wichtigen römischen Kaisern ergänzt und mit der passenden Hierarchie versehen. Im Rahmen der Arbeit wurde die Pipeline selbst mit drei neuen Funktionen modifiziert, um für die Fragestellung der Arbeit besser verwendbar zu sein. Da wir das Ergebnis von allen Münzvorderseiten-Designs zur Evaluations beider ML Ansätze brauchen, wurden mittels einer modifizierten Funktion „load_ocre_designs_obv“ lediglich die OCRE

³⁰ spaCy, <<https://spacy.io/>> (09.03.2019); Python, <<https://www.python.org/>> (09.03.2019); Jupyter Notebook, <<https://jupyter.org/>> (09.03.2019).

³¹ Klinger 2018, 14 f.; Klinger u. a. 2018, 72 f.

³² Ausschnitt aus: Klinger 2018, 13 Abb. 4; Klinger u. a. 2018, 71 Abb. 4.

³³ Klinger u. a. 2018, 72 Abb. 6.

Vorderseitenbeschreibungen aus einer anderen Tabelle geladen dann, um diese nach dem Training noch einmal durch die Pipeline zu schicken. (4) Mit „predict_single_sentence_clear“ und „predict_clear“ wurden im „model.py“ Modul zwei neue Funktionen geschrieben, damit die vom Model gefundenen Entitäten sofort als Klartext ausgegeben werden und die Ergebnisse dann in eine neue Tabelle einer relationalen Datenbank geschrieben werden können. Abschließend wurde die ganze Bedienung so vereinfacht, dass der gesamte Programmablauf vom Training des Modells bis zum Übertragen der NER Daten mit wenig Aufwand auf einem Server mit Jupyter Notebook und Zugang zur relationalen Datenbank benutzbar ist.

4.2 Bilderkennung mittels eines Convolutional Neural Network

Die zweite für die hier untersuchte Fragestellung herangezogene Arbeit war die Masterarbeit von A. Loyal zur Bilderkennung von Personenporträts und Münztypen auf antiken Münzen mittels eines *Convolutional Neural Networks (CNN)*³⁴. Aus dieser Arbeit wurde die Umsetzung der Porträterkennung weitgehend vollständig übernommen, aber für eine erleichterte Benutzbarkeit stärker modifiziert als das bei der NER Implementation der Fall war. Die Reihenfolge der einzelnen Schritte ist wiederum in Abbildung 5 zu sehen.

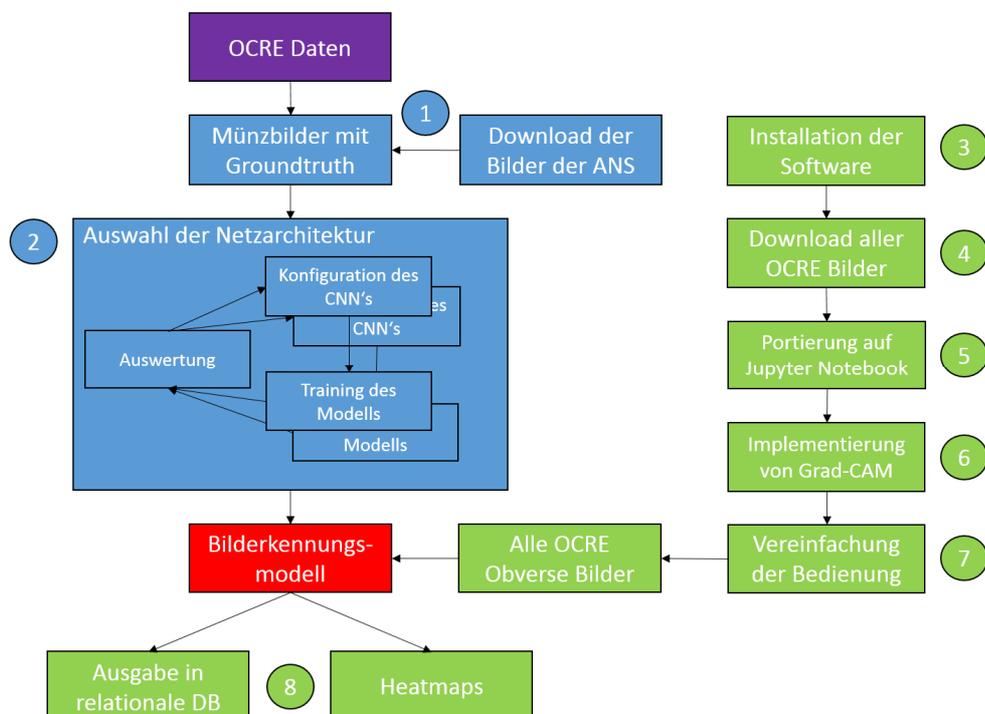


Abb. 5: Ablaufdiagramm der CNN Bilderkennung. Blau: Vorhergehende Arbeiten durch A. Loyal; Grün: Implementierungen im Rahmen dieser Arbeit.

³⁴ Loyal 2018, 63–70.

(1) Zu Beginn musste die Vorderseitenbilder der Münzen von der OCRE Webseite mittels eines Java Programms mit einer integrierten SPARQL Abfrage von M. Jostock heruntergeladen werden und in einer nach römischen Kaisern geordneten Ordnerstruktur abgelegt werden. Es wurde jedoch lediglich die *American Numismatic Society (ANS)* als Quelle verwendet³⁵. Dieser Teil der Arbeit wurde nur auf Münzvorderseiten ausgeführt, da dort im Allgemeinen die meisten Personenporträts römischer Kaiser zu finden sind. Auch hier wurden die Daten dann in Trainings- und Testmenge unterteilt. Um das Festlegen der Netzarchitektur des CNN's zu ermöglichen wurde *NVIDIA Digits* in Kombination mit dem *Deep Learning Framework Caffe* verwendet³⁶. (2) Da das Training eines CNN's sehr aufwändig ist, wurde hauptsächlich ein Rechner mit vier Grafikkarten in Verbindung mit der *NVIDIA cuDNN* Bibliothek, die die benötigte Zeit sehr stark verringert hat, benutzt³⁷. Der nächste Schritt beinhaltete die Auswahl einer Netzarchitektur. Die Autorin hat sich für die *VGG16* Architektur der Visual Geometry Group der Universität Oxford entschieden, da diese in verschiedenen Experimenten die besten Ergebnisse bei der Klassifizierung geliefert hat und außerdem dem ImageNet Datenset vortrainierte Gewichte besitzt³⁸. VGG16 besitzt 13 Convolutionsschichten und drei vollständig verbundene Schichten und ermöglicht es bis zu 1000 Klassen zu trainieren. Die Eingabebilder müssen dann auf die Größe 224 x 224 x 3 Pixel skaliert werden, um den 224 Neuronen der Eingabeschicht zu entsprechen³⁹. Anschließend hat die Autorin als Optimierungsmethode das Verfahren des *Stochastic Gradient Descent (SGD)* mit einer Lernrate von 0,001 gewählt, da diese Parameter in weiteren Tests die besten Ergebnisse erzielt haben⁴⁰. Die Bedienung des Netzes wurde über ein Python Kommandozeileninterface realisiert. Das Pythonprogramm benötigt dazu alle Konfigurationsdateien des CNN's und außerdem die Caffe Python Schnittstelle, die als Modul für Python geladen wird⁴¹. Als letztes war es notwendig den Zielordner mit den zu klassifizierenden Bildern einzutragen.

Auch an dieser Arbeit wurden einige Modifikationen vorgenommen, um die Untersuchung ihrer Ergebnisse besser ausführen zu können. (3) Um das Modell benutzen zu können ist es notwendig, die Caffe API mit Python Schnittstelle auf dem Hostrechner zu installieren. Um die Ausführungsgeschwindigkeit stark zu steigern, sollte bei vorhandener NVIDIA Grafikkarte

³⁵ Jostock 2016; American Numismatic Society, <<http://numismatics.org/>> (11.03.2019).

³⁶ NVIDIA Digits, <<https://developer.nvidia.com/digits>> (09.03.2019); Caffe, <<https://caffe.berkeleyvision.org/>> (09.03.2019); Jia u. a. 2014.

³⁷ NVIDIA cuDNN, <<https://developer.nvidia.com/cudnn>> (09.03.2019). Loyal 2018, 65–70.

³⁸ VGG16, <<https://www.kaggle.com/keras/vgg16>> (09.03.2019); Simonyan – Zisserman 2014.

³⁹ Loyal 2018, 41 f. Tab. 4.4.

⁴⁰ Loyal 2018, 45 f. 60 f.

⁴¹ Loyal 2018, 70–72.

außerdem noch cuDNN installiert werden. (4) Ferner wurde das Download-Programm von M. Jostock so angepasst, dass jetzt alle in OCRE vorhandenen Datenquellen benutzt werden⁴². Somit konnten von den über 100.000 Münzen in der Datenbank ca. 64.000 Vorderseitenbilder mit Porträts von Personen, auf deren Identifizierung das Netzwerk bereits trainiert wurde, heruntergeladen werden⁴³.

(5) Da ein Kommandozeileninterface keine einfache Bedienung ermöglicht, wurde im Rahmen dieser Arbeit der Code auf ein Jupyter Notebook portiert. Das brachte den Vorteil, dass die nötige Software auf einem Server laufen kann, während das eigentlich Programm über den Browser eines lokalen Rechners ausgeführt wird, wenn man sich mittels *Secure Shell (SSH)* mit dem Host verbindet.

(6) Aufgrund der Schwierigkeiten die CNN Bilderkennungsresultate erklärbar zu machen wurde außerdem noch die Möglichkeit hinzugefügt, sich die Bilder mit einer durch die Grad-CAM Technik erstellten Heatmap ausgeben zu lassen (siehe dazu Abschn. 3.5). Es wurden für den Benutzer zwei Optionen erstellt: 1. Die Bearbeitung, Ausgabe und Speicherung von Einzelbildern, 2. Die Bearbeitung und Speicherung der Münzbilder eines Zielordners (mit allen darin befindlichen Unterordnern). Um den Code des Programms so schlank wie möglich zu halten, wurde entschieden für die Grad-CAM Auswertung ein eigenes Notebook inklusive aller notwendigen Routinen zu erstellen. (7) Für beide Notebooks gilt, dass der Nutzer lediglich den Zielordner seiner Bilder angeben muss und dann das Notebook nur einmal komplett ausführen muss. Es wäre hier zu überlegen ein grafisches Userinterface hinzuzufügen, um die Codeebene von dem damit nicht vertrauten numismatischen Nutzer zu trennen. Abschließend wurden alle neu heruntergeladenen OCRE Bilder durch das Netzwerk propagiert und die Top-5 Ergebnisse der Klassifizierung und die dazugehörigen Wahrscheinlichkeiten werden in eine eigens erstellte Tabelle einer relationalen Datenbank eingetragen. Daneben wurde außerdem eine Reihe verschiedener Heatmaps für die folgende Auswertung angelegt (siehe Abschn. 5.3).

4.3 Übertragung der Ergebnisse ins RDF-Datenmodell

Der letzte Schritt der zur Auswertung der Ergebnisse nötigen Vorarbeit ist die Überführung der von NER Pipeline und Bilderkennung erstellten Daten von der relationalen Ebene in das RDF-

⁴² Sofern diese Quellen über eine explizite Trennung von Vorder- und Rückseitenbild einer Münze verfügen. Dies schließt beispielsweise die Bilder des British Museum in London aus, da hier die Seiten der Münzen nur als kombinierte Bilder vorliegen. Dies würde eine aufwendige, hier nicht zu leistende Trennung sämtlicher Bilder erfordern. Siehe dazu beispielsweise die Bilder des BM zu Typ: OCRE. RIC I (second edition) Augustus 9A, <[http://numismatics.org/ocre/id/ric.1\(2\).aug.9A](http://numismatics.org/ocre/id/ric.1(2).aug.9A)> (11.03.2019).

⁴³ Online Coins of the Roman Empire, <<http://numismatics.org/ocre/contributors>> (14.03.2019); Für eine Liste der von A. Loyal trainierten Personen siehe: Loyal 2018, 74 Tab. 6.1.

Datenmodell, um Abfragen mittels SPARQL zu ermöglichen (siehe Abschn. 3.6). Gleichzeitig werden die von NER und Bilderkennung gefunden Personen in eine detaillierte Hierarchie von Rollen eingebunden⁴⁴. Das Diagramm auf Abbildung 6 macht die Reihenfolge der folgenden Arbeitsschritte deutlich.

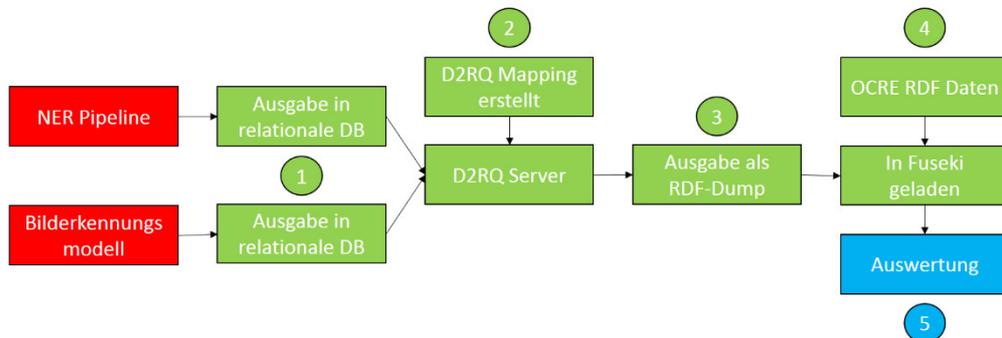


Abb. 6: Ablaufdiagramm der Übertragung der Ergebnisse in das RDF-Datenmodell

(1) Die Ergebnisse beider Ansätze werden in eigenen Tabellen einer relationalen Datenbank mit den OCRE Typen verknüpft abgespeichert. Die Übertragung der Tabellen erfolgte nun mittels der *D2RQ* Plattform⁴⁵. (2) Dafür ist eine spezielle Mappingdatei notwendig, die die Einträge der einzelnen Datensätze in RDF-Tripel, bestehend aus Subjekt, Prädikat und Objekt, umwandelt. Das RDF-Tripel besteht dann beispielsweise aus der OCRE Typ ID, welche über das Prädikat „*dcterms:hasPart*“ mit einem *rdf:Bag* Container für die Ergebnisse der NER Pipeline und einem *rdf:Seq* Container für die Resultate der Bilderkennung als sogenannte anonyme Knoten (auch leere Knoten genannt) verbunden ist. Diese Container sind mit dem „*rdf:li*“ Prädikat mit einzelnen Einträgen, die ebenfalls durch anonyme Knoten repräsentiert werden, verlinkt. Diese anonymen Knoten haben eine Verbindung zu den einzelnen Kaisern, die ebenfalls eigene Knoten bilden, über „*rdf:Subject*“. Jeder Kaiserknoten kann so mit verschiedenen Münzdesigns als NER oder Bilderkennungsergebnis verbunden sein⁴⁶.

(3) Mittels der Mapping Datei kann nun ein RDF-Dump der Ergebnisse beider Modelle angefertigt werden und auf einem SPARQL Server hochgeladen werden. In dieser Arbeit wurde dazu die *Apache Jena Fuseki* Software verwendet⁴⁷. (4) Da die Ergebnisse beider Ansätze über die OCRE Typen IDs verbunden sind, kann die Konnektivität des RDF-Models genutzt werden, um diese Ergebnisse mit den vollständigen Münzdaten der OCRE Datenbank zu verbinden.

⁴⁴ Diese Hierarchien sollen semantisch breitere Abfragen ermöglichen, siehe: Klinger u. a. 2018, 75 f. Abb. 8.

⁴⁵ D2RQ, <<http://d2rq.org/>> (11.03.2019).

⁴⁶ Siehe dazu auch: Klinger u. a. 2018, 76. Auf Klinger u. a. 2018, 76 Abb. 9 wird ein Beispiel einer solchen Verknüpfung aus dem CNT Datensatz gezeigt. Die Verbindungen für Objekte und Relationen wurden für die vorliegende Fragestellung nicht benötigt und daher weggelassen.

⁴⁷ Apache Jena Fuseki, <<https://jena.apache.org/documentation/fuseki2/>> (11.03.2019).

Dazu muss lediglich der auf der OCRE Website verfügbare RDF-Dump der Datenbank in die dieselbe Fuseki Datenbank wie der hier erstellte Ergebnis-Dump geladen werden⁴⁸. Die Ressourcen/Knoten mit gleichen URI's, die in beiden Dumps vorhanden sind, werden dann von Fuseki automatisch verschmolzen. (5) Jetzt kann diese mit neuen Daten angereicherte Datenbasis mittels SPARQL für die Auswertung abgefragt werden.

5. Evaluation der Ergebnisse

Das folgende Kapitel soll die Ergebnisse der Auswertung der Ergebnisse der *Named Entity Recognition (NER)* Pipeline und des *Convolutional Neural Networks (CNN)* zur Bilderkennung auf den *Online Coins of the Roman Empire (OCRE)* Daten vorstellen. Zu Beginn wird auf die allgemeine Performance beider Modelle mithilfe der im Forschungsgebiet des maschinellen Lernens üblichen Metriken eingegangen (siehe Abschn. 3.3). Im darauffolgenden Abschnitt werden die Ergebnisse der NER Pipeline einer genaueren Prüfung unterzogen. Dasselbe wird anschließend ebenfalls mit den Resultaten der Bilderkennung ausgeführt. Bei beiden Ansätzen werden zuerst die aufgetretenen Probleme und daraufhin die Möglichkeiten zur Verwendung diskutiert.

5.1 Performance der Modelle

In diesem Abschnitt wird zuerst ein Überblick über die Performance des NER Modells gegeben und daran anschließend die Performance der Bilderkennung besprochen.

Die NER Pipeline

Die NER Pipeline erzielte bei der Auswertung von P. Klinger auf den OCRE Designs folgende Ergebnisse:

Bei einer willkürlich gewählten Testmenge von ca. 3.500 Designs bei einer Gesamtzahl von 14.000 Designs (25%) wurden eine *Accuracy* von 90%, ein *Recall* von 96% und eine *Precision* von 99% erreicht. Da die Pipeline eigentlich für die CNT Daten, deren Anzahl an Designs deutlich geringer als die der OCRE Daten ist, entwickelt wurde, konnte bereits eine sehr gute Übertragbarkeit des Modells auf neue Daten ohne spezielle Anpassungen von der Autorin festgestellt werden⁴⁹.

⁴⁸ RDF-Dump der OCRE Datenbank, <numismatics.org/ocre/nomisma.rdf> (11.03.2019).

⁴⁹ Auf dem CNT Datensatz wurden bei einer willkürlich gewählten Testmenge eine *Accuracy* von 88%, ein *Recall* von 96% und eine *Precision* von 97% erreicht. Somit hat sich die Performance des Modells durch die größere Anzahl an Datensätzen sogar leicht verbessert. Klinger 2018, 24 f.; Klinger u. a. 2018, 73.

Seit dem Abschluss der Arbeit im Februar 2018 durch P. Klinger konnte das Modell noch weiter getestet und verbessert werden. Vor allem die Fähigkeit neue Entitäten zu finden hat sich auf den OCRE Daten als sehr erfolgreich erwiesen. Es wurden ca. 175 neue Personenentitäten von der NER Pipeline zur Personentabelle hinzugefügt, die durch unsere Domainexperten ebenfalls in die bestehende Hierarchie aufgenommen wurden. In Abbildung 7 ist ein Ausschnitt der neu gefundenen Entitäten mit Hierarchie zu sehen. Der markierte Kaiser Aurelian wurde von dem Modell in den OCRE Design Beschreibungen als eigenständige neue Entität erkannt, da er in den CNT Beschreibungen nicht vorkommt und deswegen auch nicht Teil des Trainings war.

334	Aemilian	aus OCRE	Aemilianus	(NULL)	http://nomisma.org/id/aemilian	(NULL)	Male	Human	Emperor	BarracksEmperors
336	Valerian II	aus OCRE	valerian_ii	(NULL)	http://nomisma.org/id/valerian_ii	(NULL)	Male	Human	Emperor	BarracksEmperors
337	Saloninus	aus OCRE	(NULL)	(NULL)	http://nomisma.org/id/saloninus	(NULL)	Male	Human	Emperor	BarracksEmperors
341	Youth	aus OCRE	(NULL)	(NULL)	http://www.dbis.cs.uni-frankfurt.de/cnt/id/Youth_person	(NULL)	Male	Human	(NULL)	(NULL)
344	Claudius Gothicus	aus OCRE	Claudius II, claudius_ii_gothicus	(NULL)	http://nomisma.org/id/clauidius_ii_gothicus	(NULL)	Male	Human	Emperor	BarracksEmperors
346	Aurelian	aus OCRE	(NULL)	(NULL)	http://nomisma.org/id/aurelian	(NULL)	Male	Human	Emperor	BarracksEmperors
347	Severina	aus OCRE	Ulpia Severina, ulpia_severina	(NULL)	http://nomisma.org/id/ulpia_severina	(NULL)	Female	Human	Emperor	BarracksEmperors
348	Florian	aus OCRE	Florianus	(NULL)	http://nomisma.org/id/florianus	(NULL)	Male	Human	Emperor	BarracksEmperors
349	Tacitus	aus OCRE	(NULL)	(NULL)	http://nomisma.org/id/tacitus	(NULL)	Male	Human	Emperor	BarracksEmperors
350	Probus	aus OCRE	(NULL)	(NULL)	http://nomisma.org/id/probus	(NULL)	Male	Human	Emperor	BarracksEmperors
352	Tempus	aus OCRE	(NULL)	(NULL)	http://www.dbis.cs.uni-frankfurt.de/cnt/id/Tempus_person	(NULL)	Male	Deities	Roman	Other
353	Siscia	aus OCRE	(NULL)	(NULL)	http://www.dbis.cs.uni-frankfurt.de/cnt/id/Siscia_person	(NULL)	Female	Deities	Roman	Personification
354	Savus	aus OCRE	(NULL)	(NULL)	http://www.dbis.cs.uni-frankfurt.de/cnt/id/Savus_person	(NULL)	Male	Deities	Roman	RiverGod

Abb. 7: Ausschnitt aus der Personen-Tabelle, die als Grundlage für das NER Training dient.

Diese Neufunde mit einigen wenigen manuellen Ergänzungen, welche die Anzahl der Einträge von 160 auf 362 erhöhte, ermöglichten ein neues Training auf den OCRE Daten. Dieses lieferte neue Werte mittels der in dem Programm eingebauten Auswertungsfunktionen für die hier verwendeten Metriken⁵⁰:

Die Wert für *Accuracy* konnte von 90% auf ca. 96% gesteigert werden und die *Precision* lag weiterhin bei ca. 99%. Auch der *Recall* konnte um 2% auf ca. 98% zulegen.

Damit lässt sich folgern, dass sich bei dem hier verwendeten NER Model, welches ohne Anpassungen schon sehr gut auf die OCRE Daten anwendbar war, durch eine gezielte Erweiterung der dem Training zugrundeliegenden Personenliste eine merkliche Steigerung der Klassifizierungsqualität erreichen lässt.

Jedoch muss dabei erwähnt werden, dass die Ergebnisse des Models durchweg von menschlicher Vorarbeit abhängen. Das bedeutet einerseits, dass ohne diese Vorarbeit das Model erst gar nicht tätig werden kann, da keine verwertbaren Daten, also Designbeschreibungen, vorliegen. Diese Vorarbeit ist bei Neufunden sehr aufwändig. Ein Numismatiker muss die Münzseite mit der abgebildeten Ikonographie identifizieren und eine Beschreibung des Designs in einer Datenbank erstellen. Diese Arbeit kann von wenigen Minuten bis zu einer tagelangen Recherche bei einer bisher unbekanntem Ikonographie sehr unterschiedliche Zeiträume

⁵⁰ Klinger 2018, 15 f.

annehmen. Dabei spielt ebenso der Erhaltungszustand der Münze eine Rolle. Je schlechter dieser ist, desto mehr Zeit wird in der Regel für die Identifikation gebraucht. Für die hier untersuchte Fragestellung spielt das eine eher untergeordnete Rolle, da nur die Vorderseiten von kaiserzeitlichen Münzen betrachtet werden (siehe Abschn. 2.1). Diese tragen meist ein Porträt eines Kaisers oder eines Mitglieds seiner Familie und diese sind durch zahlreiche Münzen überliefert⁵¹. Daneben liegen bereits tausende Beschreibungen verschiedenster Münzdesigns in den unterschiedlichen Datenbanken vor, auf denen die NER Pipeline tätig werden könnte.

Das zweite Problem, welches hier wichtig werden kann, ist die Frage der Datenqualität. Fehler können bei einer menschlichen Eingabe von Daten sehr leicht passieren. Dabei wird die Überprüfung der eingegebenen Daten oft sehr sporadisch ausgeführt, da das Hinzufügen neuer Datensätze als wichtiger angesehen wird als die Korrektur einiger Fehler. Dies kann im Allgemeinen zu Fehlern bei der Überprüfung wissenschaftlicher Hypothesen führen. In dem hier untersuchten Kontext kann es dazu beitragen, dass Entitäten vom NER Model nicht erkannt werden, da bereits ein Rechtschreibfehler ihren Namen zu stark verändern kann⁵². Es ist ebenso möglich, dass der fehlerhafte Datensatz zur Identifizierung einer neuen Entität durch das Model führt und man zwei Einträge für dieselbe Person in der Tabelle erhält, was dann mühsam von Hand korrigiert werden muss. In der für das Model verwendeten Personentabelle wurde daher eine Spalte für mögliche Tippfehler eines Namens hinzugefügt. Ebenso kann eine alternative Schreibweise desselben Namens zu oben angeführten Problemen führen. Auch dafür wurde eine weitere Spalte in der Tabelle angelegt (siehe Abb. 8)⁵³. Letztlich führen diese manuell ausgeführten Tätigkeiten zu einem extra Arbeitsaufwand. Daher ist hier zu überlegen, inwieweit diese Tätigkeiten automatisiert ausgeführt werden könnten.

id	name	description	alternativenames	typos	link
355	Maximian	aus OCRE	(NULL)	(NULL)	http://nomisma.org/id/maximian
356	Diodetian	aus OCRE	Diokletian	(NULL)	http://nomisma.org/id/diodetian
357	Constantius Chlorus	aus OCRE	constantius_chlorus	Constntius Chlorus	http://nomisma.org/id/constantius_chlorus
358	Quies	aus OCRE	(NULL)	(NULL)	http://www.dbis.cs.uni-frankfurt.de/cnt/id/Quies_person

Abb. 8: Einträge in der Personentabelle mit unterschiedlichen Schreibweisen und Tippfehlern

⁵¹ Howgego 1995, 69, 81, 84; Loyal 2018, 24 f.

⁵² Tolle – Wigg-Wolf in Vorbereitung.

⁵³ Es existieren nicht nur verschiedene Schreibweisen in unterschiedlichen Sprachen, sondern auch im Englischen kann beispielsweise der Kaiser *Konstantin der Große* als „Constantine the Great“ oder „Constantine I“ benannt sein. Siehe dazu auch die Auflistung unterschiedlicher Namen von Nomisma. Constantine the Great <http://nomisma.org/id/constantine_i> (13.03.2019).

Das Bilderkennungsmodell

Die Ergebnisse der Evaluation der Bilderkennung durch A. Loyal stellen sich wie folgt dar:

Von den ca. 28.000 Münzbildern wurden ca. 7000 (25%) als Testmenge willkürlich ausgewählt. Hier konnte das Netz eine *Accuracy* von 91%, einen *Recall* von 91% und eine *Precision* von ebenfalls 91% erreichen. Für diese Auswertung wurde das auf dem *ImageNet* Datensatz vortrainierte *VGG16* Netz ausgewählt, da dieses die besten Werte erreicht hat. Daneben wurde auch der F1-Score mit 90% und ein Top-1 Error von 9% ausgewiesen. Da dieses Netz auch in der Lage ist die fünf wahrscheinlichsten Ergebnisse auszugeben, wurde außerdem eine *Top-5 Accuracy* von 92% erreicht. Weiterhin wurde ein Test mit sechs Münzen mit verschiedenen Beschädigungen (starke Abnutzung, Bohrlöcher und unvollständiger Münzkörper) durchgeführt. Dabei erkannte das Netz viermal das richtige Porträt als wahrscheinlichste Klasse. Eine weitere Münze konnte innerhalb der Top-5 korrekt klassifiziert werden. Das Porträt der unvollständigen Münze wurde nicht erkannt. Die Autorin plädiert angesichts aller Ergebnisse für einen Einsatz des Netzes als Hilfsmittel der Numismatik⁵⁴.

Da aufgrund des Aufwandes kein weiteres Neutraining des Bilderkennungsnetzes durchgeführt werden konnte, bleiben die Ergebnisse der Auswertung der vorhandenen Metriken von A. Loyal grundsätzlich bestehen.

Durch die im Rahmen dieser Arbeit erfolgten Ausweitung des Datensatzes von ca. 28.000 Bildern der ANS auf ca. 64.000 Bilder verschiedener Institutionen, können zwei der oben aufgeführten Werte noch einmal überprüft werden (siehe Abschn. 4.2 mit Fußnote 42 f.). Bei ca. 52.000 Münzen wird die korrekte Klasse als wahrscheinlichstes Ergebnis vorhergesagt, was einem Top-1 Error von ca. 19% entspricht. Der Top-5 Error beträgt bei ca. 58.000 Fällen mit der richtigen Klasse in den fünf wahrscheinlichsten Ergebnissen (mit einem Treshhold von 5%) ca. 9,5%. Es fällt auf, dass der Top-1 Error bei einem Anstieg der Datenmenge um etwas mehr als das Doppelte um ca. 10% höher ausfällt. Dagegen ist der Top-5 Error nur leicht um 1,5% gestiegen (siehe Abschn. 8.2 Tabelle 1).

Das Netzwerk hat auf dem folgenden Rechner die 64.000 Münzbilder in ca. 26 Minuten bearbeitet: AMD Ryzen 5 1600, 32 GB Ram, NVIDIA Geforce GTX 1080 (8 GB). Mit der entsprechenden Hardware können somit auch große Bildermengen in kurzer Zeit durch das Modell klassifiziert werden.

Auch wenn es stärkere Abweichungen von den Ergebnissen von A. Loyal gibt, kann an dieser Stelle doch die grundsätzliche Aussage der Autorin weiterhin gelten, da die hier ermittelten

⁵⁴ Loyal 2018, 73–76 Tab. 6.1.

Werte angesichts des sehr heterogenen Materials römischer Münzen immer noch gut sind und somit die CNN basierte Bilderkennung ein nützliches Tool für Numismatiker sein kann.

5.2 Auswertung der NER Ergebnisse

Dieser Abschnitt soll zuerst die Probleme des NER Ansatzes beleuchten und anschließend die Möglichkeiten im Einsatz als Tool für Numismatiker aufzeigen.

Die folgenden Beobachtungen wurden anhand einer Abfrage mittels der SPARQL Sprache gemacht (siehe Abschn. 3.6). Diese Abfrage vergleicht die Ergebnisse der NER Pipeline direkt mit den eingetragenen Portraits der OCRE Datenbank als Groundtruth (siehe Abschn. 8. Listing 2). Zu beachten ist hierbei, dass gleiche Entitäten im OCRE RDF-Dump und dem hier erstellten NER-Dump andere URI's haben können. Das gilt vor allem für die URI's des Britischen Museums in London. Diese identifizieren die antiken Götter aufgrund des fehlenden Nomisma Eintrags. Jedoch haben sich vor kurzem diese eigentlich möglichst unveränderlichen Adressen geändert⁵⁵. Daher haben die OCRE Einträge noch die alte nicht funktionierende Adresse, während das hier verwendete Mapping auf die aktuellen URI's zugreift. Somit entstehen aus diesen und weiteren Unterschieden bei der Verlinkung in den Ergebnissen der Abfrage Einträge für unterschiedliche Entitäten, obwohl die Entitäten gleich sind.

Ein weiteres Problem bei der Verwendung des NER Modells ist die Erklärbarkeit der Ergebnisse der Klassifizierung (siehe Abschn. 3.4). Während Personen, die von dem Model im Trainingszyklus bereits gelernt wurden, durchweg mit hoher Wahrscheinlichkeit erkannt werden, gibt es auch Fälle, wo dies nicht zutrifft. Beispielsweise enthält die Beschreibung des OCRE Typs *RIC VII Sirmium 20* insgesamt drei unterschiedliche Personen: Die Kaiser *Crispus*, *Konstantin II.* und die römische Göttin *Victoria*⁵⁶. Das Modell erkennt lediglich *Crispus* und *Victoria* (siehe Abb. 9)⁵⁷. Der Kaiser *Konstantin II.* findet sich in den Ergebnissen nicht wieder, obwohl er in der für das Training maßgeblichen Personenliste eingetragen ist (siehe Abb. 10). Leider existieren für das Gebiet der *Natural Language Processing (NLP)* noch keine aussagekräftigen Tools wie Grad-CAM, die einen Beitrag zur Erklärbarkeit dieser Ergebnisse

⁵⁵ Beispielsweise lautet die alte URI für die Göttin Diana: British Museum. Diana <<http://collection.britishmuseum.org/id/person-institution/57039>>, während die neue funktionierende URI folgendermaßen aussieht: British Museum. Diana, <<https://public.researchspace.org/resource/?uri=http://collection.britishmuseum.org/id/person-institution/57039>> (16.03.2019).

⁵⁶ OCRE. RIC VII Sirmium 20, <<http://numismatics.org/ocre/id/ric.7.sir.20#obverse>> (15.03.2019).

⁵⁷ Gleichzeitig ist hier ein Fehler in der OCRE Datenbank erkennbar. Während das Modell die Göttin Victoria als Entität erkennen konnte, gibt es für die Vorderseite keinen Eintrag zu ihr. Lediglich die Rückseite weißt diesen auf, da dort zwei Victorien abgebildet sind. Siehe dazu auch Abschnitt 5.2 NER-3.

leisten könnten (siehe Abschn. 3.5). Daher fällt es schwer konkrete Gründe für das Nichterkennen von Entitäten anzugeben⁵⁸.

	design	groundTruth	nlpPerson
1	<http://numismatics.org/ocre/id/ric.7.sir.2.0#obverse>	nm.constantine_ii	nm.crispus
2	<http://numismatics.org/ocre/id/ric.7.sir.2.0#obverse>	nm.constantine_ii	cnt.Victory_person
3	<http://numismatics.org/ocre/id/ric.7.sir.2.0#obverse>	nm.constantine_ii	nm.crispus
4	<http://numismatics.org/ocre/id/ric.7.sir.2.0#obverse>	nm.constantine_ii	cnt.Victory_person
5	<http://numismatics.org/ocre/id/ric.7.sir.2.0#obverse>	nm.crispus	nm.crispus
6	<http://numismatics.org/ocre/id/ric.7.sir.2.0#obverse>	nm.crispus	cnt.Victory_person
7	<http://numismatics.org/ocre/id/ric.7.sir.2.0#obverse>	nm.crispus	nm.crispus
8	<http://numismatics.org/ocre/id/ric.7.sir.2.0#obverse>	nm.crispus	cnt.Victory_person

Abb. 9: Ergebnisse einer Abfrage in Fuseki für den Typ RIC VII Sirmium 20

445	Constans	aus OCRE
446	Constantine II	aus OCRE
447	Jovian	aus OCRE
448	Magnus Maximus	aus OCRE
449	Theodosius II	aus OCRE
450	Joannes	aus OCRE
451	Valentinian III	aus OCRE

Abb. 10: Eintrag für Konstantin II. in der Personenliste

Das Auftauchen von mehr als zwei Personen in der Beschreibung könnte eine Möglichkeit zur Erklärung dieses Verhaltens sein. Da eine solche Konstellation auf den Münzvorderseiten im OCRE Datensatz eine Ausnahme darstellt, lässt sich diese These an dieser Stelle nur schwer prüfen. Als mögliches Gegenbeispiel dazu kann jedoch der Typ RIC I (second edition) Tiberius 55 herangezogen werden, der auf der Vorderseite eine Götterversammlung an einem Tempel mit insgesamt sieben Entitäten trägt⁵⁹. Diese sieben Entitäten wurden aber von dem NER Model identifiziert. Auch der Typ RIC VII Nicomedia 39 enthält drei Entitäten in seiner Vorderseitenbeschreibung, von denen jedoch nur zwei im Porträteintrag bei OCRE angegeben sind (siehe Abb. 11)⁶⁰. Bei diesem Design gibt das NER Modell jedoch alle drei Personen aus: Konstantin der Große, Licinius und die Göttin Fortuna (siehe Abb. 12). Somit kann das Modell den OCRE Datensatz in diesem Fall sogar ergänzen. Festhalten lässt sich an dieser Stelle, dass das oben erläuterte Verhalten aufgrund noch nicht vorhandener Tools nicht erklärt werden kann und an anderer Stelle mit Daten, deren Anzahl an Beschreibungen mit mehr als zwei Personen höher ist, weiter untersucht werden muss⁶¹.

Vorderseite

Legende: DD NN CONSTANTINVS ET LICINIVS AVGG
Type: Bust of Constantine I and Licinius, laureate, draped, facing one another, together holding Fortuna
Porträt: Konstantin der Große 
Porträt: Licinius 

Abb. 11: Ausschnitt aus dem Datensatz von RIC VII Nicomedia 39 mit Vorderseitenporträts

⁵⁸ Auch hier ist eine Zunahme von möglichen Ansätzen zur Erklärbarkeit von NLP Modellen in den letzten Jahren zu beobachten, z.B.: Liu – Yin – Wang 2018. Leider konnte kein Ansatz zur Verwendung mit der spaCy API gefunden werden.

⁵⁹ OCRE. RIC I (second edition) Tiberius 55, <http://numismatics.org/ocre/id/ric.1(2).tib.55#obverse> (16.03.2019).

⁶⁰ OCRE. RIC VII Nicomedia 39, <http://numismatics.org/ocre/id/ric.7.nic.39#obverse> (16.03.2019).

⁶¹ Beispielsweise enthalten die Rückseiten der OCRE Münzen häufiger Mehrpersonendarstellungen.

design	groundTruth	nlpPerson
<http://numismatics.org/ocre/id/ric.7.nic.39#obverse>	nm:constantine_i	nm:constantine_i
<http://numismatics.org/ocre/id/ric.7.nic.39#obverse>	nm:constantine_i	nm:licinius
<http://numismatics.org/ocre/id/ric.7.nic.39#obverse>	nm:constantine_i	cnt:Fortuna_person

Abb. 12: Ausschnitt aus dem Ergebnis der Vergleichsabfrage auf dem RDF-Dump

Raum für Optimierungen der Personenliste als Grundlage der NER Pipeline sind ebenfalls noch gegeben. Die OCRE Datenbank trennt beispielsweise den Triumvirn *Octavian* von dem späteren Kaiser *Augustus*. Dagegen hat die Personenliste *Octavian* als alternativen Namen zu *Augustus* eingetragen hat. Hier wäre zu überlegen, ob man für *Octavian* ebenfalls einen eigenen Eintrag anlegt, um Abweichungen zu vermeiden⁶². Ein ähnlich gelagertes Problem gibt es bei Typen, die im Porträteintrag eine Person als vergöttlicht nennen⁶³. Da jedoch die zugehörige Beschreibung nur die nicht-vergöttlichte Person nennt, folgt hier bei der verwendeten Abfrage aufgrund der nicht vorhandenen Nomisma URI im OCRE RDF-Dump bei den Ergebnissen der Pipeline ein zusätzlicher Eintrag. Ob diese explizite Unterscheidung von Porträts in den OCRE Daten wirklich nötig ist, sollte noch einmal von den Verantwortlichen diskutiert werden.

Was sich auf den NER Daten beobachten lässt, ist, dass eine Kategorie von Fehlern, die bei der Bilderkennung einen Großteil der falschen Ergebnisse ausmachen, überhaupt nicht auftritt (siehe Abschn. 5.3): die Falschklassifizierung einer Entität. Beispielsweise wird die Entität Kaiser *Augustus* nicht als Entität Kaiser *Domitian* von dem Modell klassifiziert. Davon gibt es lediglich eine Ausnahme, die auf einem Problem der Münzbeschreibung basiert: Auf dem Typ *RIC X Valentinian III 2023* wurde durch das Modell die Entität *Aelia Eudoxia* identifiziert⁶⁴. Jedoch wurde in der Münzbeschreibung nur der Name *Eudoxia* verwendet: „*Bust of Eudoxia, pearl-diademed, draped, facing front, wearing necklace and earrings*“⁶⁵. Dagegen verweist die Inschrift der Münze auf eine Person mit ähnlichem Namen: *Licinia Eudoxia*, die Enkelin der *Aelia Eudoxia*⁶⁶. Da es aber in der Personenliste lediglich ein Eintrag namens *Eudoxia* gibt, der von dem Modell selbst hinzugefügt wurde, und dieser auf *Aelia Eudoxia* verlinkt, entsteht diese falsche Klassifizierung. Um solche Probleme beheben zu können, müsste das zugehörige

⁶² Nomisma. Augustus, <http://nomisma.org/id/augustus> (16.03.2019); Nomisma. Octavian, <http://nomisma.org/id/octavian> (16.03.2019). Hier stellt sich die Frage, in wie weit man eine Person, die verschiedene Rollen einnimmt, aufteilen sollte. Nomisma selbst empfiehlt eine Unterscheidung zwischen Octavian und Augustus, wenn diese explizit nötig ist. Eine Möglichkeit dies zu lösen, wäre das Hinzufügen eines Rollenattributes, welches die Bezeichnung des Porträts als Augustus ermöglichen würde.

⁶³ Ein Beispiel hierzu wäre: OCRE. RIC III Commodus 66, <http://numismatics.org/ocre/id/ric.3.com.661#obverse> (16.03.2019).

⁶⁴ Nomisma. Aelia Eudoxia, <http://nomisma.org/id/aelia_eudoxia> (16.03.2019).

⁶⁵ OCRE. RIC X Valentinian III 2023, <http://numismatics.org/ocre/id/ric.10.valt_iii_w.2023#obverse> (16.03.2019).

⁶⁶ Nomisma. Licinia Eudoxia, <http://nomisma.org/id/licinia_eudoxia> (16.03.2019).

Design in der Datenbank präzisiert werden, da dieses den Fehler erst ermöglicht hat. Daraufhin könnten dann in der Personenliste zwei verschiedene Einträge für *Aelia* und *Licina Eudoxia* angelegt werden, was den Fehler beheben dürfte. Ein weiteres Beispiel dafür ist auf dem Typ *RIC VI Siscia 201a* zu finden⁶⁷. Die Designbeschreibung erwähnt lediglich *Maximian*, was von der NER Pipeline als der Kaiser *Maximian* interpretiert wird⁶⁸. Jedoch ergibt sich aus der Datierung der Münze, dass diese den Kaiser *Galerius* zeigt, der ebenfalls „*Maximian*“ als Namensbestandteil trägt⁶⁹. Dieser Fehler würde sich durch eine präzisere Beschreibung des Designs ebenfalls vermeiden lassen und ist daher dem Model nicht zuzurechnen.

Ferner kann die NER Pipeline natürlich auch einen wichtigen Beitrag zu der oben schon erwähnten Datenqualität in der OCRE Datenbank leisten. Die Grundhypothese lautet, dass OCRE bereits eine sehr hohe Datenqualität besitzt. Dies gilt es nun mittels derselben Abfrage zu überprüfen, die bereits oben angewendet wurde (siehe Abschn. 8. Listing 2). Aufgrund der doch recht zahlreichen Ergebnisse dieser Abfrage und der hier nötigen händischen Überprüfung, werden an dieser Stelle einige Beispiele für die Möglichkeiten des Tools präsentiert. Die gefundenen Unstimmigkeiten sollen später an OCRE weitergeleitet werden.

a) NER-1:

Die NER Pipeline findet mehr Personen in der Münzbeschreibung als Personenporträts eingetragen sind. Dieser Fehler im OCRE Datensatz kommt nach einer Durchsicht der Abfrageergebnisse eher selten vor, da er voraussetzt, dass mehrere Personen auf der Münzvorderseite abgebildet sind und eine davon vergessen wurde oder die Identifizierung eines Einzelporträts mehrere Möglichkeiten ergibt. Der letztere Fall kann an dem Typus *RIC V Carus 150* illustriert werden⁷⁰. Die Vorderseite des Typs trägt das Porträt des Kaisers *Carinus* oder seines Vaters Kaiser *Carus*. Die unter der Herrschaft des *Carus* geprägte Münzserie nennt in ihrer Inschrift jedoch *Carinus*. Die OCRE Beschreibung zählt beide Kaiser als mögliche Identifizierungen auf. In dem zugehörigen Porträteintrag wird jedoch nur *Carinus* genannt (siehe Abb. 13)⁷¹. Die Ergebnisse des NER Modells haben in der Beschreibung beide Kaiser

⁶⁷ OCRE. RIC VI Siscia 201a, <<http://numismatics.org/ocre/id/ric.6.sis.201a#obverse>> (16.03.2019).

⁶⁸ Nomisma. Maximian, <<http://nomisma.org/id/maximian>> (16.03.2019).

⁶⁹ Nomisma. Galerius, <<http://nomisma.org/id/galerius>> (16.03.2019).

⁷⁰ OCRE. RIC V Carus 150, <<http://numismatics.org/ocre/id/ric.5.car.150>> (18.03.2019).

⁷¹ Auf der Webseite der ANS der diesem Typus zugehörigen Münze findet sich auch nur Carinus als einziger Porträteintrag: ANS. Silver Antoninianus of Carus, Lugdunum, AD 282 - AD 283 1981.91.10, <<http://numismatics.org/collection/1981.91.10>> (18.03.2019).

als mögliche Abgebildete bestimmt (siehe Abb. 14). Daher ist hier das Potential des Ansatzes zu sehen, bei einer fehlenden Porträtangabe auf diesen Sachverhalt hinzuweisen.

Vorderseite

Legende: CARINVS NOBIL CAES

Type: Bust of Carinus, radiate, draped, right, or bust of Carinus, radiate, cuirassed, right, holding spear, or bust of Carus, radiate, helmeted, cuirassed, left, holding spear in right hand and shield in left hand, or bust of Carus, radiate, cuirassed, left, holding spear in right hand and shield in left hand

Porträt: [Carinus](#)

Abb. 13: Ausschnitt aus dem Datensatz von RIC V Carus 150 mit Vorderseitenporträts

design	groundTruth	nlpPerson
< http://numismatics.org/ocre/id/ric.5.car.150#obverse >	nm:carinus	nm:carinus
< http://numismatics.org/ocre/id/ric.5.car.150#obverse >	nm:carinus	nm:carus

Abb. 14: Ausschnitt aus dem Ergebnis der Vergleichsabfrage auf dem RDF-Dump

b) NER-2:

Eine weitaus häufigere Unstimmigkeit in den Daten der Abfrage ist eine Abweichung von Porträteintrag und Person in der Beschreibung der Vorderseite des Typus. Dieser Fehler lässt sich vor allem auf eine Verwechslung der Person bei der Eintragung des Porträts oder einer doppelten Eintragung einer Person bei Autorität und Porträt zurückführen. Der erstere Fall kann an dem Typus *RIC VII Londinium 237* gezeigt werden⁷². Die dortige Vorderseitenbeschreibung nennt *Konstantin den Großen* als Porträtierten. Im Gegensatz dazu weißt der Porträteintrag seinen Sohn *Konstantin II.* aus (siehe Abb. 15). Das NER Modell kann auch hier anhand der Beschreibung den korrekten Porträteintrag finden und diesen Fehler somit sichtbar machen (siehe Abb. 16)⁷³.

Vorderseite

Legende: CONSTANTI-NVS IVN N C or CONSTANT-INVS IVN N C or CONSTANTIN-VS IVN N C

Type: Bust of Constantine I, helmeted, cuirassed, left

Porträt: [Constantin II.](#)

Abb. 15: Ausschnitt aus dem Datensatz von RIC VII Londinium 237 mit Vorderseitenporträts

design	groundTruth	nlpPerson
< http://numismatics.org/ocre/id/ric.7.lon.237#obverse >	nm:constantine_ii	nm:constantine_i

Abb. 16: Ausschnitt aus dem Ergebnis der Vergleichsabfrage auf dem RDF-Dump

Der zweite Fall dieser Kategorie lässt sich anhand des Typus *RIC V Carausius issuing for Diocletian/Maximian 48* illustrieren⁷⁴. Hier nennt die Vorderseitenbeschreibung den Kaiser

⁷² OCRE. RIC VII Londinium 237, <<http://numismatics.org/ocre/id/ric.7.lon.237#obverse>> (18.03.2019).

⁷³ Der gleiche Fehler wurde beispielsweise noch durch die Abfrage auf dem Typ OCRE. RIC VI Antioch 164b, <<http://numismatics.org/ocre/id/ric.6.anch.164b#obverse>> (18.03.2019) identifiziert.

⁷⁴ OCRE. RIC V Carausius issuing for Diocletian/Maximian 48, <http://numismatics.org/ocre/id/ric.5.cara-dio-max_her.48#obverse> (18.03.2019).

Maximian als Porträtierten. Der Porträteintrag führt jedoch den Kaiser *Carausius* an dieser Stelle an. Hier stimmen Autoritäts- und Porträtattribut *Carausius* überein. Dies lässt auf eine falsche Übernahme des Autoritätseintrags für das Porträt schließen. Dieser Fehler konnte von dem NER Modell anhand der Abfrage ebenfalls aufgedeckt werden.

c) NER-3:

Die Klasse der Unstimmigkeiten mit einem ebenfalls sehr großen Vorkommen innerhalb des Abfragedaten ist die Eintragung eines zusätzlichen Porträts, obwohl dieses nicht auf dem Vorderseitendesign vorkommt. Dieser Irrtum erfolgt meist durch die Eintragung des Porträts der Rückseite auf den Vorderseitenattributen. Auch ein unbemerkter Fehler bei der manuellen Eingabe mit der Erzeugung eines zufälligen zusätzlichen Porträts kommt vor. Der Typus *RIC V Gallienus 29* lässt sich als Beleg für den ersten Fall anführen⁷⁵. Hier existieren für die Vorderseite der Münze zwei Porträteintragungen: Kaiser *Gallienus* und die Göttin *Diana*. Jedoch beschreibt die Vorderseitendesign nur das Bildnis des Kaisers. Die Göttin ist dagegen lediglich auf der Rückseite zu finden (siehe Abb. 17). Die NER Pipeline gibt mittels der Abfrage bei diesem Typus das richtige Resultat aus: das Porträt des *Gallienus* (siehe Abb. 18). Eine Besonderheit im Rahmen dieses Fehlers soll ebenfalls erwähnt werden. Es werden zahlreiche Fälle mit der Eintragung von Gottheiten, die eigentlich der Rückseite der Münze zugehörig sind, auf Münzen des Typusklasse *RIC VI Treveri* angezeigt. Ein Beispiel für diesen Fall ist der Typus *RIC VI Treveri 598a*⁷⁶. Die große Anzahl, die die Vergleichsabfrage mit über 700 verschiedenen Typen erbringt, lässt den Schluss zu, dass bei der Übertragung der Trierer Münzen möglicherweise ein fehlerhaftes Mapping benutzt wurde.

Vorderseite
Legende: GALLIENVS AVG
Type: Head of Gallienus, radiate, right
Porträt: Gallienus 
Gottheit: Diana 

Rückseite
Legende: DIANA FELIX
Type: Diana, draped, walking right, holding bow and drawing arrow from quiver; at her feet, running hound

Abb. 17: Ausschnitt aus dem Datensatz von *RIC V Gallienus 29* mit Vorderseitenporträts

design	⚡ groundTruth	⚡ nlpPerson
<http://numismatics.org/ocre/id/ric.5.gall(2).29#obverse>	nm:gallienus	nm:gallienus
<http://numismatics.org/ocre/id/ric.5.gall(2).29#obverse>	<http://collection.britishmuseum.org/id/person-institution/57039>	nm:gallienus

Abb. 18: Ausschnitt aus dem Ergebnis der Vergleichsabfrage auf dem RDF-Dump

⁷⁵ OCRE. RIC V Gallienus 29, <http://numismatics.org/ocre/id/ric.5.gall(2).29#obverse> (18.03.2019).

⁷⁶ OCRE. RIC VI Treveri 598a, <http://numismatics.org/ocre/id/ric.6.tri.598a#obverse> (18.03.2019).

Die andere Möglichkeit der Erzeugung einer Unstimmigkeit durch Eintragung eines nicht vorhandenen Porträts kann durch den Typus *RIC II, Part 1 (second edition) Domitian 122* belegt werden⁷⁷. Dort ist neben dem richtigen Eintrag des Kaisers *Domitian* noch ein Eintrag für die Göttin *Minerva* vorhanden. Jedoch findet sich in der Beschreibung, auf den zugehörigen Münzbildern und im Ergebnis der Abfrage kein Hinweis auf die Abbildung dieser Göttin. Somit kann auch dieser Fehler durch die Abfrage auf den NER Daten aufgedeckt werden.

Fazit Insgesamt lässt sich für die NER Pipeline folgern, dass sie als eigenes Programm zur Unterstützung von Numismatikern bei Ihrer Arbeit auf verschiedenen Datenbanken sehr gut funktioniert, auch wenn teilweise etwas Extraarbeit bei der Anpassung auf neue Datensets nötig sein kann. Das Modell kann verschiedene Unstimmigkeiten in den OCRE Datensätzen aufzeigen und somit eine Korrektur ermöglichen und zur Steigerung der Datenqualität beitragen. Die mögliche Anzahl des Vorkommens aller drei Fehler ist in Abschnitt 8.2 Tabelle 2 aufgeführt⁷⁸.

Als weitere Anwendungsmöglichkeit des Modells ist im CNT Projekt beispielsweise geplant, eine für die Münzbearbeiter zusätzliche Arbeit an die NER Pipeline auszulagern: Bis jetzt wurden keine Münzseiten für die Porträteinträge der Münzen in der Datenbank angelegt. Da das NER Modell seine gefundenen Entitäten an die Münzdesigns bindet und diese mit Vorder- und Rückseiten in den Münzdatensätzen abgelegt sind, kann eine neue Tabelle mit Münze, Porträts und Münzseite des Porträts in kurzer Zeit erzeugt werden, die ansonsten bei den Bearbeitern eine tagelange Beschäftigung bei ca. 10.000 Münzen erzeugt hätte. Auch können weitere Anwendungsmöglichkeiten wie die hierarchische Suche auf semantischer Basis die Nützlichkeit dieses ML Ansatzes belegen⁷⁹.

⁷⁷ OCRE. RIC II, Part 1 (second edition) Domitian 122, <[http://numismatics.org/ocre/id/ric.2_1\(2\).dom.122#obverse](http://numismatics.org/ocre/id/ric.2_1(2).dom.122#obverse)> (18.03.2019).

⁷⁸ Diese Zahlen wurden mithilfe einer weiteren Abfrage auf den Daten der SPARQL-Vergleichsabfrage erstellt, um bestimmte Probleme auszufiltern. Die Zahlen sind daher nur vorläufige Werte, da eine händische Auswertung im Rahmen dieser Arbeit zu aufwändig wäre. So ist ein großer Teil der Typen aus Trier mit dem Fehler NER-3 nicht inkludiert. Einzelne Typen können auch zwei Fehlerklassen gleichzeitig zugeordnet sein. Beispielsweise NER-2 und NER-3: OCRE. RIC VI Treveri 689, <<http://numismatics.org/ocre/id/ric.6.tri.689#obverse>> (24.03.2019). In diesem Fall kann möglicherweise auch die Designbeschreibung falsch sein, da die Bilderkennung auch *Maximian* identifiziert.

⁷⁹ Siehe dazu: Klinger u. a. 2018, 76–78.

5.3 Auswertung der Bilderkennungsergebnisse

Diese Evaluierung soll ebenfalls zuerst die Probleme des Bilderkennungsansatzes aufzeigen und dann die möglichen Einsatzgebiete innerhalb der Numismatik besprechen. Folgende Probleme werden hier angesprochen:

- **Anzahl der Ergebnisse:** Die Top-5 Ergebnisse bringen Unstimmigkeiten hervor.
- **Unbekannte Klassen:** Nicht trainierte Klassen können nicht erkannt werden.
- **Münzzustand:** Ein schlechter Münzzustand erhöht die Fehlerquote der Identifizierung.
- **Falsch zugewiesene Münzen:** Fehler in den Trainingsdaten senken die Erkennungsrate.
- **Inschrift:** Die Bilderkennung greift anstatt auf das Porträt auf die Inschrift zurück (Testreihen: CNT und OCRE).

Diese Einsatzmöglichkeiten gibt es:

- **Fehlerklassen:** Die Fehlerklassen CNN-1 – 4 können durch die Bilderkennung identifiziert werden.

Anzahl der Ergebnisse Für die Evaluierung der Bilderkennungsergebnisse wurde eine sehr ähnliche Abfrage wie im vorhergehenden Abschnitt verwendet (siehe Abschn. 8.1. Listing 3). Bei dieser Anfrage werden die in der Datenbank eingetragenen Porträts als Groundtruth mit den Ergebnissen dieses Modells verglichen. Hier gibt jedoch zwei größere Probleme mit dem zurzeit verwendeten Ansatz: Zu jedem Design liegen aufgrund der unterschiedlichen Anzahl zugeordneter Münzen eine Vielzahl von Ergebnissen vor, da das Modell ja auf diesen Münzen arbeitet. Außerdem werden bis zu fünf Ergebnisse des Modells in dem zugrundeliegenden RDF-Dump gespeichert (siehe Abschn. 4.3). Es wird die Verbindungen zwischen einem Design- und einem Kaiserknoten bei mehrfachem Vorhandensein durch unterschiedliche Münzen, die alle einem Design bzw. Typus zugehörig sind, in den RDF-Daten auch mehrmals über die anonymen Knoten gezogen und somit können viele unterschiedliche Verlinkungen entstehen. Dies lässt sich nur über das Stichwort „distinct“ in der Abfrage wieder korrigieren, die mehrfach vorhandene Zeilen ausfiltert. Diese Bedingungen führen jedoch zu einer Vielzahl von Ergebnissen im Rahmen der vergleichenden Abfrage. Da in dieser Arbeit keine 30.000 Einträge händisch ausgewertet werden konnten, sollen an dieser Stelle wiederum Probleme und Möglichkeiten des Bilderkennungsansatzes besprochen werden.

Unbekannte Klassen Das Training des hier verwendeten CNN's fand exklusiv auf OCRE Daten statt und umfasste 62 verschiedene Klassen, die sich in den Daten der *American Numismatic Society (ANS)* finden ließen (siehe Abschn. 4.2 mit Fußnote 43). Aufgrund der

anderen Arbeitsweise eines CNN's im Vergleich zu der NER Pipeline können somit keine neuen Klassen auf noch nicht gesehenen Daten erkannt werden. Das bedeutet, dass sämtliche für die Auswertung benutzte Münzen ein Porträt aus den 62 Klassen haben sollten. Wenn man Münzen mit einer unbekanntem Klasse dem Netzwerk als Input gibt, dann werden die Ergebnisse sich dementsprechend von der Groundtruth unterscheiden.

Für diesen Zweck wurde eine Testreihe mit der Kaiserin *Crispina*, der Frau des *Commodus*, durchgeführt und dazu auch Heatmaps der Münzen erstellt. Es zeigte sich, dass die Kaiserin besonders oft als eine von drei der trainierten weiblichen Mitglieder der Kaiserhäuser identifiziert wurde: *Faustina I.*, *Faustina II.* und *Lucilla*. Hier ragt die Identifizierung als *Lucilla*, der Schwester des *Commodus* mit einem Top-1 Score (Anzahl der korrekten Identifizierungen mit größter Wahrscheinlichkeit) von 93 von 220 Münzen und einem Top-5 Score (Anzahl der korrekten Identifizierungen in den fünf wahrscheinlichsten Ergebnissen) mit 150 von 220 Münzen heraus. An zweiter Stelle der Klassifizierungen steht die Mutter des *Commodus*, *Faustina II.*, mit einem Top-1 Score von 64 / 220 und einem Top-5 Score von 129 / 220. Mit größerem Abstand folgt als drittes *Faustina I.*, die Großmutter des *Commodus*, mit folgenden Werten: Top-1 Score: 27 / 220 und Top-5 Score: 51 / 220. Damit gelingt dem Model trotz der falschen Klassifizierungen sehr gute Hinweise auf die Identität und Zeitstellung der unbekanntem Person zu geben, da die meisten Identifizierungen auf ihr sehr nahestehende Personen des Kaiserhauses fallen. Diese Personen sind dem Bilderkennungsansatz bekannt und werden somit aufgrund der Ähnlichkeit der Porträtbilder herangezogen.



Abb 19: Heatmap des Typus RIC III Commodus 281b mit Porträt der Crispina⁸⁰



Abb 20: Typus RIC III Marcus Aurelius 771 mit Porträt der Lucilla⁸¹

Der Vergleich der Porträts der *Crispina* und *Lucilla* und eine Auswertung von den erstellten Heatmaps bringt die folgenden Erkenntnisse: Es gibt eine starke Ähnlichkeit im Porträt beider

⁸⁰ ANS. Silver Denarius of Commodus, Rome, AD 178 - AD 191 1944.100.50026, <<http://numismatics.org/collection/1944.100.50026>> (19.03.2019).

⁸¹ ANS. Silver Denarius of Marcus Aurelius, Rome, AD 164 - AD 180. 1997.77.2, <<http://numismatics.org/collection/1997.77.2>> (19.03.2019).

Frauen auf den Münzen. Dies betrifft vor allem die sehr ähnlichen Frisuren⁸². Dabei orientiert sich das Netzwerk bei der Klassifizierung nach Auswertung der Heatmaps besonders an der Frisur, der Kinn-/Hals-/Gewandpartie und seltener auch dem Gesicht (siehe Abb. 19 und 20).

Während das Modell hier sehr gute Hinweise zu der abgebildeten Person bringt, gibt es jedoch auch Fälle bei denen dies nicht sichergestellt werden kann. Die Erkennung orientiert sich an wenigen Buchstaben der Münzinschrift und bringt diese dann mit einer zeitlich weit entfernten Person zusammen, die ebenfalls diese Buchstaben als Namensbestandteile besitzt. Beispielsweise wird *Domitia Longina*, eine Tante *Neros*, mit dem Kaiser *Domitian* aufgrund des gleichen Namensbestandteils „DOM“ verwechselt. Für die weitere Auswertung bedeutet dies, dass von den über 100.000 Münzen in der OCRE Datenbank aufgrund der fehlenden Klassifizierungsmöglichkeit lediglich ca. 64.000 Münzen mit Porträts trainierter Personen herangezogen werden konnten⁸³.

Münzzustand An dieser Stelle wird ein Typus mit schlechter Identifizierungsrate als Beispiel für einen Problemfall herangezogen. Es handelt sich um den Typus *RIC V Tetricus I 272* der mit 292 Münzen in der Datenbasis vertreten ist⁸⁴. Diese Münzen wurden unter der Autorität des Kaisers des gallischen Sonderreichs *Tetricus I.* mit dem Porträt seines Sohnes *Tetricus II.* herausgegeben. Während dieser Typus im Top-1 Score mit 73 / 292 falschen Identifizierungen (25 %) schon 6 % über dem hier ermittelten Top-1 Error aller Münzen liegt, beträgt der Top-5 Error mit 39 / 292 (ca. 13,4 %) ca. 4 % mehr als derjenige aller Münzen. Die Gesamtzahl der 73 Münzen mit einer falschen Top-1 Klassifizierung ragt auch unter den anderen Typen mit deutlichem Abstand hervor⁸⁵. Aufgrund dieser Zahlen wurde eine genauere Auswertung an dieser Stelle vorgenommen. *Tetricus II.* wird besonders oft als der Kaiser *Claudius II. Gothicus* mit höchster Wahrscheinlichkeit identifiziert (21 / 73). Bei den 39 Ergebnissen mit komplett falscher Top-5 ist er 29-mal vertreten. Bemerkenswert ist, dass die Verwechslung mit seinem Vater *Tetricus I.* weitaus geringer ausfällt (Top-1: 3-mal, Top-5: 7-mal). Da die Datierung des Typus von 273-274 n. Chr. nur wenige Jahre nach dem Tod des *Claudius II. Gothicus* (270 n. Chr.) ansetzt, könnte hier eine zu dieser Zeit allgemein gültige Herrscherikonographie einen Ansatz zur Erklärung bieten. Jedoch ist *Tetricus II.* auf dem

⁸² Die Frisur zeichnet sich durch sehr eine sehr eng am Kopf liegende lockige Haarsträhnen aus, die am Hinterkopf in einem abschließenden Haarknoten/-schnecke münden.

⁸³ OCRE, <<http://numismatics.org/ocre/>> (19.03.2019).

⁸⁴ OCRE. RIC V Tetricus I 272, <http://numismatics.org/ocre/id/ric.5.tet_i.272> (20.03.2019).

⁸⁵ Diese Zahlen wurden mit Abfragen auf der relationalen Datenbank ermittelt, da eine Unterscheidung von Ergebnissen mit höchster Wahrscheinlichkeit aufgrund der Reihenfolge innerhalb des rdf:Seq Containers noch nicht verwirklicht werden konnte.

untersuchten Typus als jüngerer, bartloser Mann abgebildet, der die Strahlenkrone trägt. *Claudius II. Gothicus* wird auf seinen Münzen allgemein als bärtiger, etwas älterer Mann porträtiert, der ebenfalls eine Strahlenkrone trägt. Die Heatmaps zeigen eine starke Fixierung des Modells auf diese Strahlenkrone, wenn *Tetricus II.* als *Claudius II. Gothicus* identifiziert wird. Das Bilderkennungsnetzwerk scheint also vor allem diesen Gegenstand als ein Merkmal des *Claudius II. Gothicus* anzusehen (siehe Abb. 21).



Abb 21: Links: Typus RIC V *Claudius Gothicus* 192⁸⁶; Mitte: Typus RIC V *Tetricus I* 272⁸⁷; Rechts: Heatmaps der Münze in der Mitte, die als *Claudius II. Gothicus* identifiziert wurde

Was bei weiterer Durchsicht der Daten der komplett falsch bestimmten Münzen dieses Typus auffällt, ist dass 37 der 39 Münzen aus der Sammlung des Münzkabinetts der Heinrich Heine Universität Düsseldorf stammen. Insgesamt gibt es 135 Münzen dieses Typs aus Düsseldorf in der Datenbank und 98 davon wurden in den Top-5 richtig bestimmt. Alle Münzen stammen aus einem Schatzfund bei Brauweiler⁸⁸. Die Bilder dieser Münzen aus diesem Schatzfund zeigen, dass die Meisten sehr starke Zerstörungsspuren zeigen (siehe Abb 21. Mitte). Ebenso sind die Oberflächen dieser Münzen sehr stark abgerieben, so dass das Porträt nur noch teilweise zu erkennen ist. Die Oberflächen der korrekt als *Tetricus II.* erkannten Münzen aus Düsseldorf zeigen teilweise auch größere Zerstörungsspuren, aber meist weniger Abrieb. Hier verraten die Heatmaps, dass das bartlose Kinn des *Tetricus II.* wesentlich öfter als Identifizierungsmerkmal herangezogen wird. Damit lässt sich insgesamt festhalten, dass das Bilderkennungsmodell, sobald die Münze größere Zerstörungsspuren oder starken Abrieb aufweist, eine größere Anzahl an falschen Klassifizierungen hervorbringt. Trotz allem liegen die Klassifizierungen nicht völlig daneben, da immer noch 32 der 39 Münzen mit komplett falscher Top-5 als Porträts von

⁸⁶ ANS. Silver Antoninianus of Claudius II Gothicus, Siscia, AD 268 - AD 270 1944.100.32521, <<http://numismatics.org/collection/1944.100.32521>> (20.03.2019).

⁸⁷ Münzkabinetts der Heinrich-Heine-Universität Düsseldorf. b6385.z972, <<http://www3.hhu.de/muenzkatalog/ikmk/object?id=ID619>> (20.03.2019).

⁸⁸ Ziegler 1983.

Personen des 3. Jhs. n. Chr. identifiziert werden. Damit kann der Ansatz immer noch als Hinweisgeber für den Numismatiker dienen.

Falsch zugewiesene Münzen Ein weiteres Problem der CNN Bilderkennung ist das Training einer Klasse mit nicht zugehörigen Bildern. Es handelt sich hierbei um die OCRE Münzen aus Trier, die einen zusätzlichen Porträteintrag mit Göttern, die auf der Rückseite dargestellt wurden, auf der Vorderseite tragen (siehe Abschn. 5.2.c). Bei Durchsicht der Trainingsbilder, die mit der Groundtruth „*Genius*“ gekennzeichnet waren, lässt sich beobachten, dass diese zum größten Teil aus Porträtmünzen von Kaisern bestehen. 12 von 142 Trainingsbilder zeigen wirklich das Bildnis des *Genius*, der Rest besteht aus Münzen mit Porträts des *Maximian*, *Diokletian* und weiteren Kaisern⁸⁹. Damit muss das Netzwerk ein sehr heterogenes Bild der Merkmale dieser Gottheit entwickelt haben. Dies zeigt sich auch in den vorliegenden Daten. Bei 55 von 351 Münzen mit der Groundtruth „*Genius*“ wurde die Gottheit als wahrscheinlichstes Ergebnis erkannt. Bei den Top-5 ist der Wert mit 178 von 351 besser. Typen die auch in den Trainingsdaten vorhanden waren wie *RIC VIII Constantinople 22* weisen jedoch mit einem Top-1 Wert von 10 /12 und Top-5 Wert von 11 / 12 wesentlich bessere Erkennungsraten auf. Dieses Problem zeigt abermals, wie wichtig die Datenqualität in einer öffentlich zugänglichen Datenbank wie OCRE ist. Erstaunlich ist dabei, dass trotz der zum größten Teil falschen Trainingsdaten die Bilderkennung immer noch in der Lage ist, die Gottheit wenigstens in den Top-5 mit ca. 50% Erkennungswahrscheinlichkeit zu identifizieren. Jedoch könnte sich dieser Fehler negativ auf die Erkennungsrate der falsch als zur „*Genius*“ Klasse zugehörig gekennzeichneten Kaiser auswirken. Bei einem Neutraining des Netzwerks ist unbedingt darauf zu achten diesen Fehler zu korrigieren.

Inschrift Ein nicht zu unterschätzendes Problem des Bilderkennungsmodells ist die Inschrift der Trainingsmünzen. Das Problem äußert sich bei der Durchsicht der erstellten Heatmaps der Münzen durch den Fokus des Modells auf die Inschrift als Grundlage der Klassifizierung.

Testreihe CNT Das Problem wurde entdeckt als versuchsweise 532 Münzbilder der *Corpus Nummorum Thracorum (CNT)* Datenbank mit Porträteinträgen von Kaisern, die dem Bilderkennungsnetzwerk durch das Training bekannt waren, als Eingabedaten für die Bilderkennung genutzt wurden⁹⁰. Die Ergebnisse der Top-1 und Top-5 Scores auf diesen

⁸⁹ Loyal 2018, 52 f.

⁹⁰ Es sollte ein möglichst breites Spektrum an Kaisern abgebildet werden. Folgende Kaiserporträts waren vertreten: *Domitian*, *Gallienus*, *Hadrian*; *Marc Aurel*, *Nero*, *Phillipus Arabs*, *Septimius Severus*, *Tiberius Trajan* und *Valerian*.

Münzen waren überraschend. Von 523 Porträtmünzen wurde bei lediglich 10 Münzen die korrekte Klassifizierung als wahrscheinlichstes Ergebnis gefunden. Nicht viel besser fiel der Top-5 Wert dieses Tests aus: 25 / 532. Von Interesse ist in diesem Zusammenhang auch die durchschnittliche Wahrscheinlichkeit des besten Ergebnisses. Diese beträgt hier ca. 30%, während der Wert für die OCRE Ergebnisse mit ca. 89 % wesentlich höher ausfällt. Diese Werte deuteten somit auf ein größeres Problem des Bilderkennungsansatzes mit diesen neuen Münzen hin, die einige Unterscheide zu denen der OCRE Datenbank aufweisen: Bei den CNT Münzen, die aus der antiken Landschaft Thrakien stammen, handelt es sich um lokal produzierte Serien der thrakischen Städte und Stämme, die ebenfalls ein Kaiserbildnis tragen können. Dagegen decken die Münzbestände aus der OCRE Datenbasis den weitaus größeren Raum des römischen Reiches ab. Ferner tragen die thrakischen Münzen meist eine griechische Inschrift, die auf OCRE Münzen nicht vorkommt⁹¹. Weiterhin sind viele CNT Münzen nur als Gipsabdruck fotografiert worden, während die OCRE Münzen im Original als Foto vorliegen.

Gipsabdrücke Bei der Durchsicht der erstellten Heatmaps der hier verwendeten CNT Münzen wurde auch die Gipsabdrücke von den Münz-Domain Experten mögliche Erklärung der großen Fehleranzahl vorgebracht. Diese sind lediglich eine Kopie der eigentlichen Münze und können somit zusätzliche Fehler enthalten. Sie haben außerdem einen völlig anderen Farbton als die Originale aus Metall (siehe Abb. 22. Links). Zu den Fehlern dieser Kopien zählen vor allem die weniger scharfen Kanten/Umrisse des Münzreliefs, die vielleicht durch eine Bildbearbeitung besser hervorgehoben werden könnten.

Griechische Inschrift Auch die griechische Inschrift könnte hier eine Rolle spielen: Die Heatmaps zeigen, dass das Bilderkennungsmodell oft die Inschrift, wenn diese gut erkennbar ist, als Grundlage seiner Klassifikation benutzt und das eigentliche Porträt auslässt. Dies verwundert angesichts der dem Netzwerk eigentlich unbekanntem griechischen Inschriften. Hier wird der Eindruck erweckt, dass die Buchstaben genutzt werden, die sich nicht vom lateinischen Alphabet unterscheiden. Beispielsweise scheinen die Buchstaben mit einem Fokus auf den beiden O's (Omikron) auf der thrakischen Münze in Abbildung 22 (Rechts) als Grundlage für die Klassifikation als *Constantinus Chlorus* verwendet zu werden, während das eigentliche Porträt nichts zur Bestimmung beiträgt.

⁹¹ Selbst die neun Münzen, die in der OCRE Datenbank einer thrakischen Münzstätte zugewiesen sind, besitzen eine lateinische Inschrift. Auch ist diese Lokalisierung bei allen Exemplaren als unsicher gekennzeichnet. OCRE, <http://numismatics.org/ocre/results?q=mint_facet%3A%22Thrakien%22> (21.03.2019).



Abb 22: Links: Münze Bizye (Byzantion) CNT_1187; Rechts: Heatmap derselben Münze⁹²

Testreihe OCRE Um diese These zu prüfen wurden zwei weitere Serien von Heatmaps mit jeweils ca. 500 Münzen der OCRE Datenbank angelegt. Die erste Serie enthält nur Münzen mit Augustusporträts verschiedener Typen. Die zweite Serie besteht dagegen aus verschiedenen Typen mit dem Porträt des *Antoninus Pius*.

Die Augustusserie besteht aus 500 Münzen, die alle mit der höchsten Wahrscheinlichkeit als *Augustus* von dem Bilderkennungsmodell klassifiziert wurden. Von diesen Münzen haben 110 Heatmaps einen größeren Fokus des Netzwerks auf die Inschrift neben dem Porträt erkennen lassen (siehe Abb. 23. Links). Eine kleinere Münzreihe innerhalb dieser Serie trägt keine Inschrift auf der Münzvorderseite. Hier ist durchgehend zu erkennen, dass sich das Modell in diesem Fall ausschließlich auf das Porträt konzentriert (siehe Abb. 23. Mitte).

Münzen des *Antoninus Pius* machen den Befund noch deutlicher. Von 509 Münzen wurden ca. 288 mit einer deutlichen Einbeziehung der Inschrift als *Antoninus Pius* klassifiziert. Dabei gibt es viele Exemplare, bei denen der Fokus des Modells ausschließlich auf der Inschrift liegt (siehe Abb. 23. Rechts). Daraus lässt sich insgesamt schließen, dass die Münzen der CNT Datenbank durch ihre Unterschiede zu denen der OCRE Datenbasis ein eigenes Netzwerk mit angepasstem Training benötigen, um die niedrigen Werte der Klassifizierung beheben zu können. In der Arbeit von A. Loyal war ursprünglich angedacht, dass das Bilderkennungsnetzwerk die Kaiser anhand ihres Porträts erkennen sollte⁹³. Da sich herausgestellt hat, dass das Modell die Inschriften der Münzen sehr oft für seine Identifizierungen heranzieht, sollte überlegt werden, ob man diese vor einem erneuten Training aus den Bildern entfernt, um eine Fokussierung auf das Porträt zu erreichen⁹⁴.

⁹² CNT. Bizye (Byzantion) CNT_1187, <<https://www.corpus-nummorum.eu/coins.php?id=1187>> (21.02.2019).

⁹³ Loyal 2018. 29 f.

⁹⁴ Die Bilderkennung folgt hier den Numismatikern, da für diese die Inschrift bei der Identifizierung einer Porträtmünze ebenfalls eine wichtige Rolle spielt.



Abb 23: Links: Heatmap des Typus RIC I (second edition) Augustus 42A⁹⁵; Mitte: Heatmap des Typus RIC I (second edition) Augustus 36A⁹⁶; Rechts: Heatmap des Typus RIC III Antoninus Pius 54B⁹⁷

Fehlerklassen Natürlich lässt sich der hier untersuchte Bilderkennungsansatz ebenfalls zur Steigerung der Datenqualität der OCRE Datenbank verwenden. Grundsätzlich können dieselben drei Klassen von Unstimmigkeiten in den Daten aufgedeckt werden wie bei der NER Pipeline (siehe Abschn. 5.2 a-c). Dabei ist jedoch eine umfangreichere Auswertung wegen der Top-5 Ergebnisse des Modells zu leisten. Gute Hinweise können hier eine Abwesenheit von Übereinstimmungen zwischen der von OCRE stammenden Groundtruth und den bis zu fünf Ergebnissen des Modells für eine Münze liefern. Es muss hier jedoch auf besondere Fälle wie das oben beschriebene Problem mit dem Typus *RIC V Tetricus I 272* beachtet werden. Im Rahmen dieser Arbeit sollen die Ergebnisse der mit der NER Pipeline gefundenen Unstimmigkeiten mit den Ergebnissen der Bilderkennung verglichen werden.

a) NER-1 vs. CNN-1:

Der Typus *RIC V Carus 150* (siehe Abschn. 5.2 a), welcher nur mit einem Bild in den Daten vertreten ist, liefert als Ergebnis des hier untersuchten Modells mit 84 % Wahrscheinlichkeit das Ergebnis *Carinus*⁹⁸. Jedoch muss in diesem Zusammenhang erwähnt werden, dass der Kaiser *Carus* nicht in der Liste der trainierten Personen auftaucht. Daher sollte auch hier wieder ein Blick auf die Heatmap geworfen werden. Diese zeigt wiederum deutlich, dass die Identifikation hier völlig über die Inschrift gelaufen ist, die den Namen *Carinus* enthält. Das Bilderkennungsmodell kann also in diesem Fall keine Hilfe sein, da das Inschriftenproblem hier völlig dominiert und noch zu wenige Klassen trainiert sind.

⁹⁵ ANS. Silver Denarius of Augustus, Caesaraugusta, 20 BC - 19 BC 1948.19.1030, <<http://numismatics.org/collection/1948.19.1030>> (20.03.2019).

⁹⁶ ANS. Silver Denarius of Augustus, Caesaraugusta, 19 BC - 18 BC. 1937.158.408 <<http://numismatics.org/collection/1937.158.408>> (20.03.2019).

⁹⁷ ANS. Silver Denarius of Antoninus Pius, Rome, AD 139 1956.127.352, <<http://numismatics.org/collection/1956.127.352>> (20.03.2019).

⁹⁸ OCRE. RIC V Carus 150, <<http://numismatics.org/ocre/id/ric.5.car.150>> (22.03.2019).

b) **NER-2 vs. CNN-2:**

Der vermeintlich falsche Portraiteintrag des Typus *RIC VII Londinium 237* (siehe Abschn. 5.2 b) *Konstantin II.* wird durch das Bilderkennungsmodell bestätigt⁹⁹. Mit rund 74 % wird *Konstantin II.* als wahrscheinlichstes Ergebnis ausgegeben. Der Blick auf die Heatmap der einzig verfügbaren Münze mit Bild dieses Typus zeigt, dass das Modell hier den Stirn-/Helmbereich und einen Teil der Rüstung als wichtigste Kriterien herangezogen hat. Die Inschrift wird mit lediglich dem Buchstaben „C“ wahrgenommen (siehe Abb. 24. Links).



Abb 24: Links: Heatmap des Typus *RIC VII Londinium 237*¹⁰⁰; Rechts: Zum Typus *RIC VIII Arelate 31* zugewiesene Münze des Fitzwilliam Museums¹⁰¹

Für den zweiten vom NER Modell aufgedeckten Fehler des Typus *RIC V Carausius issuing for Diocletian/Maximian 48* findet sich leider keine zugewiesene Münze in der Datenbasis und kann daher hier nicht behandelt werden¹⁰². An dieser Stelle bietet die Bilderkennung einen weiteren Blickwinkel auf diese Unstimmigkeit des besprochenen Typus und stellt somit sogar die oben gezogene Schlussfolgerung des falschen Portraiteintrags in Zweifel. Gleichzeitig zeigt sich auch, dass natürlich Münzen mit Bildern in der Datenbasis vorhanden sein müssen, damit das Modell seine Arbeit verrichten kann.

c) **NER-3 vs. CNN-3:**

Die fehlerhafte Eintragung einer auf der Rückseite eines Typus eingetragenen Gottheit kann auch durch das Bilderkennungsmodell aufgedeckt werden. Die Göttin *Diana* als Portraiteintrag des Typus *RIC V Gallienus 29* wird in den Ergebnissen des Modells nicht gefunden (siehe Abschn. 5.2 c)¹⁰³. Hier sollte ebenfalls erwähnt werden, dass Diana nicht Teil der trainierten

⁹⁹ OCRE. *RIC VII Londinium 237*, <<http://numismatics.org/ocre/id/ric.7.lon.237#obverse>> (22.03.2019).

¹⁰⁰ Fitzwilliam Museum. CM.NW.235-R, <<http://webapps.fitzmuseum.cam.ac.uk/explorer/index.php?oid=149289>> (22.03.2019).

¹⁰¹ Fitzwilliam Museum. CM.1902-1950, <<http://webapps.fitzmuseum.cam.ac.uk/explorer/index.php?oid=185075>> (22.03.2019).

¹⁰² OCRE. *RIC V Carausius issuing for Diocletian/Maximian 48*, <http://numismatics.org/ocre/id/ric.5.cara-dio-max_her.48#obverse> (22.03.2019).

¹⁰³ OCRE. *RIC V Gallienus 29*, <[http://numismatics.org/ocre/id/ric.5.gall\(2\).29#obverse](http://numismatics.org/ocre/id/ric.5.gall(2).29#obverse)> (22.03.2019).

Personen ist. Es sind jedoch auch alle anderen Ergebnisse des Modells für die zwei vorhandenen Münzen des Typus mit den Einträgen *Claudius II. Gothicus* und *Genius* fehlerhaft. Daher kann die Bilderkennung bei diesem Typus wenig zur Entdeckung von Unstimmigkeiten beitragen. Für die in diesem Fehlerfall erwähnten Trierer Münzen kommt an dieser Stelle wiederum das oben erwähnte Trainingsproblem zum Tragen: Der Typus *RIC VI Treveri 598a* wurde durch den Portraiteintrag *Genius* zur Training der *Genius*-Klasse herangezogen¹⁰⁴. Daher ist es auch erklärbar, dass die Bilderkennungsergebnisse für die beiden Münzen dieses Typus zweimal *Genius* mit entweder ca. 20 % oder 5 % auf dem dritten Rang beinhalten. Somit kann die Bilderkennung nur durch die wahrscheinlichste Klasse bei diesem Fehler weiterhelfen. Dies belegt die Wichtigkeit von korrekten Trainingsdaten noch einmal deutlich. Bei dem abschließenden dritten Beispiel dieser Fehlerart, kann der Ansatz etwas besser sein Potential zeigen. Der Typus *RIC II, Part 1 (second edition) Domitian 122* mit dem fehlerhaften *Minerva* Portraiteintrag wird bei zwei vorhanden Münzen durch das Bilderkennungsmodell mit zweimal *Domitian* als wahrscheinlichstes Ergebnis richtig klassifiziert¹⁰⁵. Die Göttin *Minerva* ist zwar nicht unter den trainierten Klassen, aber bei einem ausgeweiteten Neutraining des Netzwerks mit fehlerfreien Klassendaten, sollte keine Verwechslung entstehen. Abschließend lässt sich zu dieser Fehlerklasse sagen, dass die fehlenden Klassen und die Klasse mit den falschen Trainingsdaten den größten Einfluss auf das Potential des Ansatzes nehmen. Daher sollte diese Klasse nach Beseitigung dieser Missstände wesentlich bessererkennbar sein.

d) CNN-4:

Ein gänzlich anderer Fehler, der die Münzzuweisung zu einem Münztypus betrifft, kann exklusiv durch das Bilderkennungsmodell identifiziert werden. Der Typus *RIC VIII Arelate 31* wird mit dem Porträt der Stadtgottheit *Konstantinopolis* beschrieben und besitzt auch denselben Portraiteintrag (siehe Abb. 24. Rechts)¹⁰⁶. Die einzig zugewiesene Münze des Typus befindet sich in der Sammlung des Fitzwilliam Museum in Cambridge und hat auf deren Webseite eine andere Vorderseitenbeschreibung als die des Typus¹⁰⁷. Das Bilderkennungsmodell gibt hier vier Ergebnisse aus: *Licinus*, *Julian*, *Licinus II* und *Maximian*. Da die Klasse *Konstantinopolis* zu den trainierten Klassen gehört, wäre ein Auftauchen im Ergebnis nicht unwahrscheinlich. Doch bei dem Blick auf die Webseite mit den Daten der Münze zeigt sich, dass das Modell mit ca.

¹⁰⁴ OCRE. RIC VI Treveri 598b, <<http://numismatics.org/ocre/id/ric.6.tri.598a#obverse>> (22.03.2019).

¹⁰⁵ OCRE. RIC II, Part 1 (second edition) Domitian 122, <[http://numismatics.org/ocre/id/ric.2_1\(2\).dom.122#obverse](http://numismatics.org/ocre/id/ric.2_1(2).dom.122#obverse)> (18.03.2019).

¹⁰⁶ OCRE. RIC VIII Arelate 31, <<http://numismatics.org/ocre/id/ric.8.ar.31>> (22.03.2019).

¹⁰⁷ Auch die anderen Daten unterscheiden sich auffällig: Fitzwilliam Museum. CM.1902-1950, <<http://webapps.fitzmuseum.cam.ac.uk/explorer/index.php?oid=185075>> (22.03.2019).

36 % Wahrscheinlichkeit das richtige Porträt an zweiter Stelle gefunden hat: Kaiser *Julian*. Aus dieser Unstimmigkeit lässt sich nun zeigen, dass die zugewiesene Münze keinerlei Merkmal des Typus trägt und somit falsch verlinkt wurde. Diesen Fehler hätte das NER Modell nicht finden können, da hier Beschreibung und Porträteintrag übereinstimmen und nur ein Blick auf die Münze selbst diese Unstimmigkeiten aufdecken kann.

Fazit Das Fazit der Auswertung zum Bilderkennungsmodell lautet, dass hier durchaus viel Potential zu einem zuverlässigen Hilfstool für Numismatiker und einem Ansatz zur Steigerung der Datenqualität einer Münzdatenbank vorhanden ist. Jedoch ist es dafür unbedingt nötig noch einmal ein Neutraining des Netzwerks anzusetzen, um Fehler zu korrigieren und die Fixierung auf die Inschrift der Münzen zu beseitigen. Eine andere Möglichkeit wäre die Konzentration auf eine Typbestimmung von Münzen durch Inschrift und Bild gleichzeitig.

6. Resümee und Ausblick

Resümee Der Untersuchungsgegenstand dieser Arbeit waren die von P. Klinger und A. Loyal erstellten, auf maschinellen Lernmethoden basierenden Modelle der *Named Entity Recognition* Pipeline und der Bilderkennung mittels *Convolutional Neural Networks*. In diesem Rahmen wurden die Möglichkeiten und Probleme der bereits trainierten Modelle als Hilfstoole für die Numismatik auf antiken römischen Münzdaten der *Online Coins of the Roman Empire* und dem *Corpus Nummorum Thracorum* getestet.

Um die Ergebnisse der beiden Ansätze besser vergleichen zu können, mussten die bisherigen Implementationen modifiziert werden. Für das NER Modell wurden nur wenige Funktionen und eine veränderte Ausgabe der Ergebnisse in eine relationale Datenbank hinzugefügt. Das Bilderkennungsmodell erhielt dagegen mehrere Modifikationen. Nach der Installation der notwendigen Software auf einem neuen Rechner wurde der Quellcode in das *Jupyter Notebook* Format portiert und für eine leichtere Bedienung angepasst. Außerdem wurde eine neue Grad-CAM Implementation hinzugefügt, um Heatmaps von Münzbildern zur Auswertung erzeugen zu können. Mittels des Tools von M. Jostock zum Download der OCRE Bilder wurden jetzt insgesamt ca. 64.000 Münzen als Eingabedaten für das Netzwerk beschafft. Eine Ausgabe in eine relationale Datenbank wurde ebenso hinzugefügt. Der letzte Schritt der Implementierung war die Überführung der Ergebnisse beider Ansätze in das *Resource Description Framework* – Format, damit auf einem neu erzeugten RDF-Dump die Auswertung durch verschiedene SPARQL Abfragen erfolgen konnte.

Die Auswertung der Metriken für maschinelle Lernverfahren ergab für das NER Modell eine messbare Steigerung der Performance aufgrund des größeren OCRE Datensatzes und der Verfeinerung der zugrundeliegenden Personenliste. Das Bilderkennungsmodell zeigte dagegen bei zwei Metriken ein deutliches Absinken der Performance.

Als nächstes erfolgte eine detaillierte Evaluation der NER Ergebnisse. Dabei wurde festgestellt, dass OCRE veraltete URI's des British Museums benutzt und daher Unstimmigkeiten bei einer vergleichenden Abfrage auf den NER Ergebnissen und der eingetragenen Groundtruth entstehen. Ein weiteres Problem ist bei diesem Ansatz die mangelnde Erklärbarkeit der Ergebnisse, da eigentlich trainierte Entitäten nicht immer erkannt werden. Von Vorteil ist dagegen, dass die Falschklassifizierung einer Entität lediglich durch eine zu wenig detaillierte Designbeschreibung entstehen kann. Dieser Ansatz ist zur Steigerung der Datenqualität der OCRE Datenbasis geeignet. Es wurden drei Unstimmigkeitsklassen erstellt, die aufgedeckt werden können: NER-1: Ergänzung fehlender Porträtangaben, NER-2: Berichtigung verwechselter Porträteinträge und NER-3: Entfernung zusätzlicher, falscher Porträts. Im Rahmen von NER-3 konnte beispielsweise ein grundsätzliches Problem mit Porträteintragungen von Münzen aus Trier aufgedeckt werden. Es wurden daneben bereits über 70 Unstimmigkeiten mit der NER Pipeline in den OCRE Daten entdeckt. Dies zeigt, dass dieser Ansatz ein sehr gutes Hilfstool für Numismatiker sein kann.

Bei der Auswertung der Bilderkennungsergebnisse konnten einige größere Probleme erkannt werden. Durch die Eintragung von bis zu fünf Ergebnissen pro Münze (Top-5) und mehreren Münzen pro Typ wird durch eine Vergleichsabfrage eine Vielzahl händisch auszuwertender Datensätze erzeugt. Zurzeit ist die Klassifizierungsmöglichkeit des Netzwerkes noch eingeschränkt, da erst 62 Personen aus OCRE trainiert wurden. Die Erkennung für untrainierte Klassen kann aber gute Hinweise für die Identifizierung geben, wenn die Inschrift der Münze nicht bei der Klassifizierung einbezogen wurde. Ein großes Problem ist der Zustand der Münzen. Hier zeigte sich, dass ein starker Abrieb der Münzoberfläche oder fehlende Teile des Münzkörpers den Ansatz an seine Grenzen bringen und somit schlechte Erkennungsraten produziert werden. Besondere Aufmerksamkeit muss auch dem Trainingsdatensatz des Netzwerkes zuteilwerden. Münzen mit nicht zu einer Klasse gehörenden Porträts in den Trainingsdaten können die Erkennungsraten dieser Klasse stark absenken. Das wohl größte Problem der Bilderkennung ist aber die Fokussierung des Netzwerkes auf die Inschrift der Münzen anstatt auf das Porträt. Dies wurde anhand einer Testreihe auf Münzbildern der *Corpus Nummorum Thracorum* Datenbank mittels Heatmaps aufgedeckt und durch zwei weitere

OCRE Testreihen bestätigt. Diese Heatmaps zeigten, dass der Ansatz sehr oft die Inschrift einer Münze als Klassifizierungsgrundlage benutzte, was dem Ziel der Porträterkennung zuwiderlief.

Es wurde aber ebenso festgestellt, dass die Bilderkennung ebenso mit vier möglichen Fehlerarten zur Steigerung der Datenqualität beitragen kann. Die ersten drei Fehlerklassen entsprechen den drei Klassen der NER Pipeline. Eine eigene vierte Klasse kann dagegen nur exklusiv durch die Bilderkennung aufgedeckt werden: Die Falschzuweisung einer Münze zu einem Typus mit einem anderen Porträt. Somit besitzt auch dieser Ansatz ein großes Potential als Hilfstool zur Steigerung der Datenqualität einer Datenbank und bei der Identifizierung neuer Fundmünzen, wenn man die genannten Probleme berücksichtigt und korrigieren kann.

Ausblick Als Ausblick für die **NER Pipeline** ist zu sagen, dass einer Verwendung als Hilfstool in der Numismatik nichts entgegensteht. Dazu zählt die oben schon erwähnte Seitenbestimmung von Porträts und die semantische Suche auf der CNT Datenbank. Auch aufgrund der guten Übertragungsfähigkeit auf andere Datenbanken ist eine weitreichende Verwendung zur Steigerung der Datenqualität denkbar. Lediglich dem Problem des Nichterkennens von Personen sollte mit anderen Daten noch einmal nachgegangen werden.

Dagegen gibt bei dem **Bilderkennungsnetz** noch eine Reihe von Möglichkeiten zur Optimierung der Porträterkennung. Um das Inschriftenproblem zu lösen sollten die Trainingsdaten einer vorherigen Bearbeitung mit Trennung von Porträt und Inschrift unterzogen werden. Dann wäre es möglich, zwei getrennte neuronale Netze auf jeweils einem Bestandteil zu trainieren und somit bessere Ergebnisse zu erhalten. Mit diesen Ergebnissen wäre auch eine Typenerkennung möglich. Wenn diese Aufteilung in dem Umfang nicht geleistet werden kann, wäre ein weiterer Ansatz die mit Inschrift und Bild gleichzeitig arbeitende Typenerkennung, die A. Loyal schon in ihrer Arbeit behandelt hat, zu fokussieren. Für das Neutraining des Netzwerkes sollte auch die auf 64.000 Bilder ausgeweitete Datenbasis zum Einsatz kommen. Dies hätte den Vorteil, dass die vorhanden 62 Klassen besser trainiert werden könnten und neue Klassen, die in diesen Bildern enthalten sind, mitaufgenommen werden könnten. Auch ein Test mit einer anderen API wie Tensorflow kann eine Möglichkeit zur Verbesserung der Ergebnisse und zur Verringerung der Trainingszeit auf der vorhandenen Hardware sein.

Eine andere zu untersuchende Möglichkeit ist die Korrektur der Ergebnisse des einen Ansatzes durch den Anderen. Beide Modelle können dieselben Unstimmigkeiten (NER-/CNN-1-3) entweder aus der Perspektive der Designbeschreibung oder dem Münzbild aufdecken. Die

Bildererkennung kann sogar einen weiteren Fehler identifizieren (CNN-4). Die folgenden Punkte sind dabei denkbar und in Abb. 25 visualisiert: a) Es ist möglich, dass beispielsweise eine schlecht erhaltene Münze von der Bildererkennung falsch identifiziert wird, aber Groundtruth und NER Pipeline übereinstimmen, also ein Fehler der Bildererkennung sehr wahrscheinlich ist. b) Wenn NER und Bildererkennung das gleiche Ergebnis vorweisen, dann ist ein Fehler der Groundtruth anzunehmen. c) Bei einer Übereinstimmung von Bildererkennung und Groundtruth wird wahrscheinlich ein Fehler bei der NER Pipeline vorliegen (CNN-4).

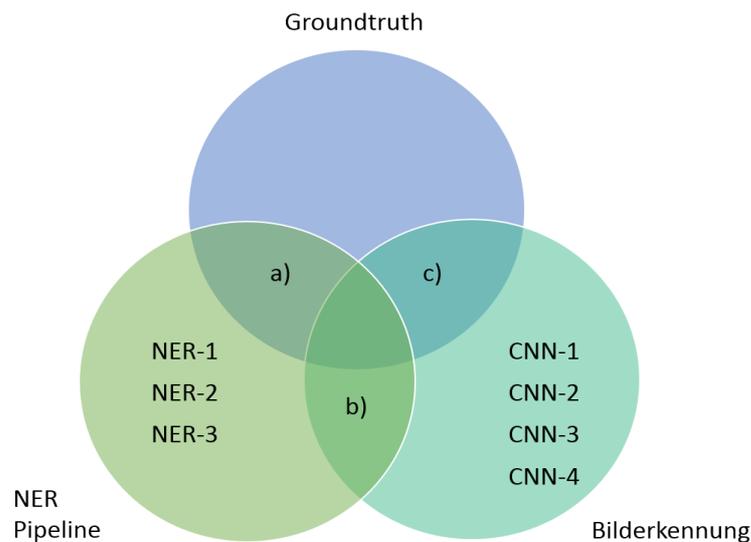


Abb. 25: Venn-Diagramm mit Groundtruth, NER Pipeline, Bildererkennung und Fehlerklassen

Ergebnisse Ein direkter Vergleich der Ergebnisse beider Ansätze und der in der Datenbank eingetragenen Groundtruth hat somit das Potential die Unstimmigkeiten, die in den Datensätzen einer numismatischen Datenbank vorhanden sind und ansonsten nur zufällig gefunden werden könnten, aktiv aufzudecken. Jedoch ist für die Entscheidung, welches Problem vorliegt, noch eine zeitaufwändige händische Auswertung und vorhandenes Domain-Wissen nötig.

Diese Arbeit und ihre Ergebnisse haben gezeigt, dass die beiden auf maschinellen Lernverfahren basierenden Methoden der NER Pipeline und der CNN basierten Bildererkennung als Unterstützungsprogramme im Bereich der Numismatik viele Möglichkeiten zur Steigerung der Datenqualität und bei Behebung der Schwierigkeiten der Bildererkennung sogar zur automatischen Identifikation neuer Fundmünzen bieten. Eine hohe Datenqualität der OCRE Datenbank konnte bis auf die Ausnahme der Münzen aus dem antiken Trier bestätigt werden. Die gefundenen Unstimmigkeiten sollen an OCRE zur Korrektur weitergeleitet werden. Es erscheint daher angemessen, beide Modelle als Forschungsgegenstand weiterzuerfolgen und in Zusammenarbeit mit Numismatikern weiterzuentwickeln.

7. Literaturverzeichnis

Bird u. a. 2009

S. Bird – E. Klein – E. Loper, Natural Language Processing with Python. – Analyzing Text with the Natural Language Toolkit (Sebastopol 2009).

Gruber 2018

E. Gruber, Linked Open Data and Hellenistic numismatics, in: S. Glenn – F. Duyrat – A. Meadows (Hrsg.), Alexander the Great. A Linked Open World (Pessac 2018) 17–34.

Goodfellow – Bengio – Courville 2016

I. Goodfellow – Y. Bengio – A. Courville, Deep Learning (Cambridge 2016).

Heath 2018

S. Heath, SPARQL as a first step for querying and transforming numismatic data. Examples from nomisma.org, in: S. Glenn – F. Duyrat – A. Meadows (Hrsg.), Alexander the Great. A Linked Open World (Pessac 2018) 35–54.

Howgego 1995

C. Howgego, Ancient History from Coins (London 1995).

Jia u. a. 2014

Y. Jia – E. Shelhamer – J. Donahue – S. Karayev – J. Long – R. Girshick – S. Guadarrama – T. Darrell, Caffe. Convolutional Architecture for Fast Feature Embedding, <<https://arxiv.org/pdf/1408.5093.pdf>> (11.03.2019).

Jostock 2016

M. Jostock. Automatische Münzerkennung mit OpenCV, Bachelorarbeit (Frankfurt 2016).

Klinger 2018

P. Klinger, Natural Language Processing to enable semantic search on numismatic descriptions, Bachelorarbeit (Frankfurt 2018).

Klinger u. a. 2018

P. Klinger – S. Gampe – K. Tolle – U. Peter, Semantic search based on Natural Language Processing - A numismatic example, Journal of Ancient History and Archaeology 5.3, 2018, 68–79.

Liu – Yin – Wang 2018

H. Liu – Q. Yin – W. Y. Wang, Towards Explainable NLP. A Generative Explanation Framework for Text Classification, <<https://arxiv.org/pdf/1811.00196.pdf>> (15.03.2019).

Loyal 2018

A. Loyal, Maschinelles Lernen angewendet auf Bilder antiker Münzen. Masterarbeit (Frankfurt 2018).

Petkovic u. a. 2018

D. Petkovic – R. Altman – M. Wong – A. Vigil, Improving the explainability of Random Forest classifier – user centered approach, in: R. Altman – A. Dunker – L. Hunter (Hrsg), Biocomputing 2018. Proceedings of the Pacific Symposium. Internationales Symposium Biocomputing Kohala Coast, Hawaii 3–7.1.2018, 204–215.

Selvaraju u. a. 2017

R. R. Selvaraju – M. Cogswell – A. Das – R. Vedantam – D. Parikh – D. Batra, Grad-CAM. Visual Explanations from Deep Networks via Gradient-based Localization, <<https://arxiv.org/pdf/1610.02391.pdf>> (04.03.2019).

Simonyan – Zisserman 2014

K. Simonyan – A. Zisserman, Very deep convolutional networks for largescale image recognition, <<https://arxiv.org/pdf/1409.1556.pdf>> (15.03.2019).

Tolle – Wigg-Wolf in Vorbereitung

K. Tolle – D. Wigg-Wolf, Improving Data Quality by Rules – A Numismatic example, Computer Applications and Quantitative Methods in Archaeology (CAA). Internationale Fachkonferenz vom 14.–16. März 2017 in der Georgia State University, Atlanta/USA (in Vorbereitung).

Ziegler 1983

R. Ziegler, der Schatzfund von Brauweiler (Bonn 1983).

8. Anhang

8.1 SPARQL Abfragen

Listing 1: Verwendete Abkürzungsprefixe

PREFIX rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>

prefix rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>

prefix nif: <<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>>

PREFIX dcterms: <<http://purl.org/dc/terms/>>

PREFIX foaf: <<http://xmlns.com/foaf/0.1/>>

PREFIX nm: <<http://nomisma.org/id/>>

PREFIX nmo: <<http://nomisma.org/ontology#>>

PREFIX org: <<http://www.w3.org/ns/org#>>

PREFIX skos: <<http://www.w3.org/2004/02/skos/core#>>

PREFIX dcterms: <<http://purl.org/dc/terms/>>

PREFIX dct: <<http://purl.org/dc/terms/>>

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX db: <http://dbpedia.org/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
prefix cnt: <http://www.dbis.cs.uni-frankfurt.de/cnt/id/>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
```

Listing 2: Abfrage zur Anzeige aller sich unterscheidender Einträge von OCRE und der NER Pipeline

```
SELECT ?design ?groundTruth ?nlpPerson ?cnnPerson
WHERE {
  ?design nmo:hasPortrait ?groundTruth.
  ?design dcterms:hasPart ?bag1.
  ?bag1 a rdf:Bag.
  ?bag1 rdf:li ?entry1.
  ?entry1 rdf:subject ?subj1.
  ?subj1 skos:exactMatch ?nlpPerson.
  FILTER (?groundTruth != ?nlpPerson).
}
```

Listing 3: Abfrage zur Anzeige aller sich unterscheidender Einträge von OCRE und dem Bilderkennungsmodell

```
SELECT distinct ?design ?groundTruth ?cnnPerson
WHERE {
  ?design nmo:hasPortrait ?groundTruth.
  ?design dcterms:hasPart ?bag2.
  ?bag2 a rdf:Seq.
  ?bag2 rdf:li ?entry2.
  ?entry2 rdf:subject ?subj2.
  ?subj2 skos:exactMatch ?cnnPerson.
  FILTER (?groundTruth != ?cnnPerson).
}
```

Listing 4: Abfrage zur Anzeige aller Ergebnisse von der NER Pipeline und dem Bilderkennungsmodell

```
SELECT ?design ?groundTruth ?nlpPerson ?cnnPerson
WHERE {
  ?design nmo:hasPortrait ?groundTruth.
  ?design dcterms:hasPart ?bag1.
  ?design dcterms:hasPart ?bag2.
  ?bag1 a rdf:Bag.
  ?bag2 a rdf:Seq.
  ?bag1 rdf:li ?entry1.
  ?bag2 rdf:li ?entry2.
  ?entry1 rdf:subject ?subj1.
  ?entry2 rdf:subject ?subj2.
  ?subj1 skos:exactMatch ?nlpPerson.
  ?subj2 skos:exactMatch ?cnnPerson.
}
```

8.2 Tabellen

Tabelle 1: Kennzahlen der Bilderkennungsauswertung

Auswertung	Anzahl Bilder	Top-1 Error	Top-5 Error
Loyal 2018	28.000	9%	8%
Neu	64.000	19%	9,50%

Tabelle 2: Mögliche Anzahl der vom NER-Model gefundenen Fehler

Fehlerart	NER-1	NER-2	NER-3	Gesamt
Mögliche Fehler	18	24	37	79

Diese Zahlen wurden mithilfe einer weiteren Abfrage auf den Daten der SPARQL-Vergleichsabfrage erstellt, um bestimmte Probleme auszufiltern. Die Zahlen sind daher nur vorläufige Werte, da eine händische Auswertung im Rahmen dieser Arbeit zu aufwändig wäre. So ist ein großer Teil der Typen aus Trier mit dem Fehler NER-3 nicht inkludiert. Einzelne Typen können auch zwei Fehlerklassen gleichzeitig zugeordnet sein.