

Friends with Benefits: Wie Deep-Learning basierte Bildanalyse und kulturhistorische Heraldik voneinander profitieren

Hiltmann, Torsten

hiltmann@uni-muenster.de

Humboldt-Universität zu Berlin, Deutschland

Thiele, Sebastian

s.thiele@uni-muenster.de

Westfälische Wilhelms Universität Münster, Deutschland

Risse, Benjamin

b.risse@uni-muenster.de

Westfälische Wilhelms Universität Münster, Deutschland

Wappen und ihre Überlieferung als kulturhistorische Herausforderung

Wappen zählen zu den am häufigsten gebrauchten visuellen Zeichen und Kommunikationsträgern des Mittelalters und der Frühen Neuzeit. Allein für das mittelalterliche Westeuropa sind über eine Million unterschiedliche Wappen bekannt (Pastoureaux 2018, 42). Von fast allen sozialen Schichten gebraucht, konnten diese in den verschiedensten Techniken auf den unterschiedlichsten Materialien dargestellt werden. Dabei waren die Wappen nicht nur einfache Identifikationsmarken für ihre Besitzer, sondern Träger komplexer Kommunikationsakte, die Identität, Besitz und Herrschaft ebenso ausdrücken konnten wie Parteizugehörigkeit, (behauptete) Herkunft und Verwandtschaft, oder auch politische Konzepte, Schutz, Ehre, Schande usw. (Paravicini 1998). Sie bildeten damit ein zentrales Kommunikationsmittel, dessen Analyse umfangreiche Einblicke in die vormoderne Kultur und Gesellschaft erlaubt. Dass dies in der bisherigen Forschung jedoch kaum geschah, mag an drei Gründen liegen, die mit den traditionellen Methoden der Geschichts- und Kulturwissenschaften nur schwer zu überwinden sind: die schiere Menge der Überlieferung, die Diversität der unterschiedlichen Gebrauchskontexte, sowie die Komplexität der Wappen als Medien an sich (Hiltmann 2019).

Während für die Frage der Komplexität und der Analyse umfangreicher heraldischer Daten bereits erste digitale Lösungsansätze entwickelt wurden (Hiltmann/Riechert

2019), sind die Möglichkeiten zur Erfassung der breiten und diversen Überlieferung noch ungeklärt. Dabei ist dies die grundlegende Voraussetzung dafür, die Entwicklung der Wappen und ihres Gebrauchs über Zeiten, Räume und soziale Gruppen hinweg nachvollziehen zu können.

Wappen als Bilddomäne für Deep Learning Algorithmen

Eine mögliche Lösungsstrategie zur systematischen Erschließung visuell transportierter Wappendaten kann die maschinelle Bildanalyse liefern. Insbesondere aktuelle Durchbrüche im Bereich des Deep Learnings konnten erstaunliche Ergebnisse in der visuellen Objekterkennung erzielen. Neben der hohen Performanz und Genauigkeit dieser lernenden Algorithmen ermöglicht deren Training via Backpropagation auch die Disambiguierung komplexer Strukturen in großen Datensätzen (LeCun et al. 2015).

Dem kommt entgegen, dass sich die visuelle Struktur der Wappen im Unterschied zu den meisten anderen Bilddomänen relativ einfach formalisieren lässt. Denn bei den Wappen handelt es sich nicht um Bilder im klassischen Sinne, sondern um abstrakte Codes aus Formen und Farben. Für die Darstellung und Wiedererkennung eines Wappens ist es wichtig, dass die mit den einzelnen Komponenten verbundenen Konzepte (z.B. Löwe, Kreuz, Lilie) erkennbar sind. Wie diese jedoch konkret dargestellt wurden, kann von Abbildung zu Abbildung variieren. Das heißt, dass es für die Darstellungen des gleichen Wappens einen breiten Spielraum gab.

Für die automatische Bildanalyse stellen die Wappen daher eine neue und besonders interessante Bilddomäne dar. Auf maschinellem Lernen basierende Bildanalysealgorithmen sind bislang besonders sensitiv für Texturmerkmale (Geirhos et al. 2019). Da die Textur eines Wappens jedoch häufig durch das Trägermedium und die jeweilige Darstellungstechnik determiniert ist, kann diese hier im Allgemeinen vernachlässigt werden. Stattdessen sind Wappen primär durch geometrische Primitive sowie deren Anordnung charakterisiert. Damit bietet sich hier für die maschinelle Bildanalyse die Möglichkeit, die Funktionsweise der Algorithmen für verschiedene Bildabstraktionen (z.B. Geometrie vs. Textur; Form vs. Farbe, etc.) zu untersuchen und diese entsprechend weiterzuentwickeln.

Zielstellung

Der vorliegende Beitrag beschreibt die konkrete Entwicklung neuer digitalen Ressourcen und Methoden für die kulturhistorische Forschung. Dabei macht er zugleich deutlich, wie die computergestützte Analyse von Wappendarstellungen die Entwicklung und das Verständnis von Bildanalysetechniken erweitern kann. Tatsächlich ist es erst das enge Ineinandergreifen

kulturhistorischer und bildanalytischer Kompetenzen und Fragestellungen, das für beide beteiligten Domänen neue und hoch innovative Potentiale für die weitere Forschung eröffnet (Abbildung 1).



Abbildung 1: Interaktion zwischen Heraldik und Bildanalyse. Farbige Pfeile zeigen die interdisziplinären Abhängigkeiten, graue Boxen listen die Vorteile für die jeweilige Disziplinen auf. Die in dieser Arbeit adressierten Aspekte sind mit einem Haken markiert.

In einem ersten Schritt soll es dabei darum gehen, Wappendarstellungen auf unterschiedlichen Medien automatisch erfassen und verzeichnen zu lassen. Dabei konzentriert sich der vorliegende Beitrag zunächst auf die Detektion heraldischer Abbildungen in mittelalterlichen und frühneuzeitlichen Handschriften unter Rückgriff auf die Methoden der Deep Learning basierten Bildanalyse.

Konkrete Umsetzung

Erstellung des Datensatzes

Der Einsatz von Deep Learning Algorithmen auf Wappendarstellungen erfordert eine hinreichend große Bilddatenmenge, welche eine spezifische, für diese Algorithmen verständliche Struktur aufweisen muss. Da bis heute keine solche Datengrundlage verfügbar ist (Sustek 2018), bestand unser erster Schritt in der Erstellung einer entsprechenden Datenbank. Diese speist sich dabei aus 34 einschlägigen Handschriften aus der Bibliothèque nationale de France (BnF), der Bibliothèque municipale de Bourges und der Bayerischen Staatsbibliothek (BSB) München, wie sie von den betreffenden Bibliotheken über deren Internetportale als Digitalisate bereitgestellt werden (siehe zukünftig: <http://digitalheraldry.org>).

Die ausgewählten Handschriften stammen aus dem 14. bis 17. Jahrhundert und wurden so ausgewählt, dass sie die mögliche Bandbreite von Wappendarstellungen in Handschriften abbilden (Hofman 2019), von einfachen Einzeldarstellungen (Besitzeinträge) und ihrer Verwendung in bildlichen Darstellungen über ungeordnete Skizzensammlungen (Epitaphiensammlungen, Genealogien) bis hin zu geordneten Wappenbüchern. Dabei sind sowohl Wappendarstellungen in Schildform enthalten, mit unterschiedlichen Neigungen, Skalierungen und Ausgestaltungsformen, als auch deren freie Darstellung auf Kleidung und Fahnen. Die zugrundegelegten Textgenres reichen dabei von unterschiedlichen Traktaten und Wissenssammlungen über literarische und

historiographische bis hin religiösen und liturgischen Texten.

Wie Abbildung 2A zeigt, wurden die Wappen auf den entsprechenden Digitalisaten der Handschriftenseiten mittels einer Boundingbox markiert, wozu das Labeling Tool LabelImg (Tzutalin 2015) verwendet wurde. Insgesamt beinhaltet die Datenbank damit 7.182 Wappendarstellungen, welche sich auf 1.568 Seiten verteilen.

Data Augmentation mit Style Transfer zur Texturabstraktion

Da Wappen vornehmlich durch die Geometrie und nicht durch die Textur determiniert sind, haben wir in einem zweiten Schritt die Datenbank um visuell augmentierte Varianten der Wappenbilder ergänzt (*Data Augmentation*). Insbesondere haben wir hier auf die Methoden des *Style Transfer* zurückgegriffen, mit denen der Stil eines Bildes verändert wird während der eigentliche Bildinhalt gleich bleibt. Anfang 2019 konnten Geirhos und Kollegen zeigen, dass mittels Style Transfer trainierte neuronale Netze mehr von der Textur eines Bildes abstrahieren und stärker Geometrien und Formen erlernen (Geirhos et al. 2019). In unserem Fall wurde der *Style Transfer* zur Vergrößerung der Datenbank mit Hilfe von AdaIN (Huang 2017) realisiert. Als Referenz dienten hierbei die Bilder der Datenbank "Painter by Numbers" von Kaggle.com, welche ca. 80.000 Gemälde verschiedener Künstler umfasst. Die Variationsbreite der enthaltenen Stile verhindert, dass das neuronale Netz, welches mit diesen Daten trainiert wird, den Stil eines spezifischen Künstlers lernt. Jedes Bild der Datenbank wurde in einen zufällig ausgewählten Stil transferiert und damit die Gesamtmenge der Trainingsdaten verdoppelt. Ein Beispiel für einen solchen *Style Transfer* ist in Abbildung 2 gegeben.



Abbildung 2: Data Augmentation mit Hilfe von Style Transfer. (A) Originalbild aus der Handschrift (Wappenposition via roter Box eingezeichnet). (B) Augmentiertes, synthetisch erstelltes Bild. (C), (D) Vergrößertes Wappen aus (A) und (B).

Deep Learning basierte Wappendetektion

Anschließend wurden zwei aktuelle Verfahren zur Objektdetektion auf dieser augmentierten Datenbank trainiert und analysiert. In unserer Studie haben wir YOLOv3 (Redmon 2018) und RetinaNet (Lin 2017) untersucht. In einer ersten qualitativen Analyse hat insbesondere das RetinaNet eine sehr hohe Genauigkeit erzielt, weshalb wir uns im Folgenden auf RetinaNet beschränken. Die Architektur dieses 2017 publizierten Deep Learning Objekt Detektor Modells ist in Abbildung 3 dargestellt. Dabei handelt es sich um einen einstufigen Detektor, der auf einem sogenannten *Feature Pyramid Network* aufbaut, das wiederum auf einem ResNet (He 2015) als Feature Extractor basiert. Dieser Aufbau ermöglicht es, Vorhersagen auf verschiedenen Auflösungsstufen eines Eingabebildes durchzuführen. Der Output dieser verschiedenen Stufen wird jeweils durch ein Klassifikations- und ein Regressionssubnetz zur Vorhersage der Boundingboxen verarbeitet. Außerhalb des Trainings werden diese Ausgaben zusätzlich noch durch *Non-Maximum-Suppression* gefiltert.

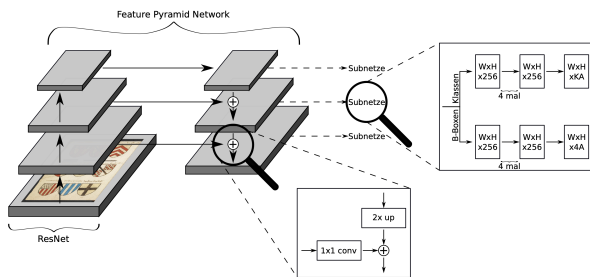


Abbildung 3: RetinaNet Architektur. Mittels Feature Pyramid Network werden relevante Merkmale auf verschiedenen Skalierungsstufen extrahiert. Anschließend werden via Subnetze Vorhersagen für Boundingboxen (B-Boxen) errechnet (angelehnt an (Lin 2017)).

Vorläufige Ergebnisse

Zum Training wurde die Datenbank in ein Trainings- und Testdatensatz aufgeteilt. Der Trainingsdatensatz stammte aus 20 Handschriften, enthielt insgesamt 1.090 Seiten mit Wappendarstellungen und wurde mittels Style Transfer augmentiert. Die übrigen 14 Handschriften im Testdatensatz enthielten 478 Seiten. Nach dem Training des RetinaNets auf dem Trainingsdatensatz erzielten wir eine Average Precision von 0,80 und einen F1 Score von 0,78 auf den Testdaten (Flach 2015).



Abbildung 4: Ergebnisse des RetinaNet Wappendetektors. Grüne Boxen zeigen manuell gesetzte Wappenpositionen, die roten Boxen die vom Wappendetektor erkannten Regionen.

In der Analyse der Ergebnisse ist festzustellen, dass Wappen sowohl über verschiedene Skalierungsstufen (vgl. Abb. 4 C und D) als auch über verschiedene Wappenstile und damit über verschiedene Texturen hinweg (vgl. Abb. 4 D und E) zuverlässig erkannt wurden. Dabei können Wappen auch in unübersichtlichen Szenen (z.B. Abb. 4A unten) detektiert werden.

Ferner zeigt sich, dass Wappen in Schildform sehr gut erkannt werden (z.B. Abb. 4C) , während Wappen auf Kleidung und Fahnen deutlich schlechter abschneiden und maßgeblich den Fehler der Average Precision und des F1 Scores beeinflussen (Abb. 4B). Dies lässt sich möglicherweise damit erklären, dass die Trainingsdaten zwar auch Wappen auf Kleidung und Fahnen enthalten, diese jedoch im Vergleich zu der deutlich weiter verbreiteten Schildform stark unterrepräsentiert sind. Um diese Fehlerquelle zu beheben, soll in einem nächsten Schritt ein Klassifikator trainiert werden, der nicht nur die Position und Größe eines möglichen Wappens erkennt, sondern auch entscheidet, ob es sich um ein heraldisches Schild, eine heraldische Fahne oder um heraldische Kleidung handelt.

Einordnung der Ergebnisse

Das Projekt macht nachvollziehbar, wie im Rahmen des maschinellen Lernens mit ganz unterschiedlichen Spielräumen umgegangen werden kann. Während durch den *Style Transfer* in den Trainingsdaten der Erkennungsraum von der Textur gelöst und auf die Geometrie der Wappen umgeleitet wurde, muss er hinsichtlich der konkreten Operationalisierung dessen, was hier als Wappen verstanden wird, für die Maschine wiederum deutlich eingeschränkt bzw. präzisiert werden. Für die maschinelle Bildanalyse ergeben sich damit neue methodische Potenziale, da

die Komplexität von Wappendarstellungen zwischen populären Trainingsdatensätzen wie handgeschriebenen Ziffern (z.B. MNIST) und allgemeinen Fotos (z.B. COCO) liegen und die es an weiteren Daten (z.B. Wappendarstellungen auf anderen Materialien) zu präzisieren gilt. Neben der Notwendigkeit der konzeptionellen Schärfung der gebrauchten Konzepte zeichnet sich mit dem Projekt für die kulturhistorische Heraldik zugleich die Möglichkeit ab, die bereits umfassend digitalisierten Handschriftenbestände von Bibliotheken wie der BnF und der BSB unter Hinzunahme der jeweils hinterlegten Metadaten erstmals umfassend auf Fragen der Verbreitung und Verwendung von Wappendarstellungen in Handschriften zu untersuchen. Ein Ansatz, der im weiteren Projektverlauf dann auch auf Wappendarstellungen in anderen Medien (Siegel, Münzen, Museumsobjekte, Wandmalereien etc.) übertragbar ist.

Paravicini, Werner (1998): "Gruppe und Person. Repräsentation durch Wappen im späteren Mittelalter" in: Oexle, Otto Gerhard / Hülsen-Esch, Andrea von (eds.): *Die Repräsentation der Gruppen. Texte - Bilder - Objekte*, Göttingen: Vandenhoeck&Ruprecht 327-389.

Pastoureau, Michel (2018): *L'art héraldique au Moyen Âge*, Paris:Seuil.

Redmon, Joseph / Farhadi, Ali (2018): "Yolov3: An incremental improvement." in: *arXiv preprint arXiv:1804.02767*.

Sustek, Martin / Vidensky, Frantisek / Zboril, Frantisek / Zboril, Frantisek (2018): "Family Coat of Arms and Armorial Achievement Classification." in: *International Conference on Intelligent Systems Design and Applications: 577-586*.

Tzutalin (2015): "LabelImg" in: Git code <https://github.com/tzutalin/labelImg>.

Bibliographie

Flach, Peter / Kull, Meelis (2015): "Precision-recall-gain curves: PR analysis done right" in: *Advances in Neural Information Processing Systems* 28: 838-846.

Geirhos, Robert / Rubisch, Patricia / Michaelis, Claudio / Bethge, Matthias / Wichmann, Felix / Brendel, Wieland (2019): "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness" in: *Proceedings of the International Conference on Learning Representations (ICLR)*.

He, Kaiming / Zhang, Xiangyu / Ren, Shaoqing / Sun, Jian (2016) „Deep Residual Learning for Image Recognition“ in: *Proceedings of the IEEE conference on computer vision and pattern recognition*: 770-778.

Hiltmann, Torsten (2019): "Zwischen Grundwissenschaft, Kulturgeschichte und digitalen Methoden. Zum aktuellen Stand der Heraldik" in: *Archiv für Diplomatik* 65: 287-319.

Hiltmann, Torsten / Riechert, Thomas (2020): "Digital Heraldry. The State of the Art and New Approaches Based on Semantic Web Technologies" in: Christelle Balouzat-Loubet (ed.), *L'édition en ligne de documents d'archives médiévaux*, Turnhout: Brepols 102-125 [im Druck].

Hofman, Elmar (2019): *Armorial in medieval manuscripts. Collections of coats of arms as means of communication and historical sources in France and the Holy Roman Empire (13th – early 16th centuries)*, PhD University of Münster.

Huang, Xun / Belongie, Serge (2017): „Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization“ in: *Proceedings of the IEEE International Conference on Computer Vision*: 1501-1510.

LeCun, Yann / Bengio, Yoshua / Hinton, Geoffrey (2015): "Deep learning" in: *Nature* 521 (7553): 436-444.

Lin, Tsung-Yi / Goyal, Priya / Girshick, Ross / He, Kaiming / Dollár, Piotr (2017): "Focal loss for dense object detection" in: *Proceedings of the IEEE international conference on computer vision*: 2980-2988.