Contextualizing Bandera: Ein Distant Watching Ansatz

Bermeitinger, Bernhard
Bernhard.Bermeitinger@uni-passau.de
Lehrstuhl für Informatik mit Schwerpunkt Digital Libraries and Web Information Systems, Universität Passau, Deutschland

Howanitz, Gernot
Gernot.Howanitz@uni-passau.De
Lehrstuhl für Slavische Literaturen und Kulturen, Universität Passau, Deutschland

Radisch, Erik
Erik.Radisch@uni-passau.de
Lehrstuhl für Digital Humanities, Universität Passau, Deutschland

Einleitung

Zahlreiche geisteswissenschaftliche Projekte im Kontext der Digital Humanities sind textlastig. Ein Grund dafür ist das breite Spektrum an etablierten Verfahren, das für solche Fragestellungen zur Verfügung steht. Aus der Perspektive der Kulturwissenschaften ergibt sich hier ein *desideratum*; schließlich widmen sich diese der (menschlichen) Kultur in ihrer ganzen Bandbreite und decken kulturelle Äußerungen im weitesten Sinne ab, die unterschiedlichste Medien, physische Artefakte und performative Handlungen mit einschließen. Zwar ist es eingeschränkt möglich, kulturelle Phänomene zu transkribieren, also in textuelle Form zu bringen, was aber kaum automatisierbar ist und Informationsverluste birgt. Native Ansätze, welche auf jeweils spezifische Eigenschaften des zu untersuchenden Phänomens eingehen, erscheinen deshalb vielversprechend.

Neben Texten spielen Bilder in zahlreichen kulturellen Zusammenhängen eine tragende Rolle, so auch im Internet: Bilder und Videos werden kopiert, bearbeitet, geteilt und damit zu sogenannten Memen (Shifman 2013). Dabei entsteht eine große Zahl an Bildern bzw. Videos, die sich aufgrund ihrer schieren Masse einem traditionellen *Close Reading* entzieht. Gleichzeitig liegen die Bilder und Videos digitalisiert vor und sind zum Teil durch die verwendete Web-Plattform (z. B. YouTube) mit Metadaten annotiert. Aus diesen Gründen sind Bilder-Meme und virale Videos prädestiniert für den Einsatz quantitativer Verfahren.

Die hier vorgestellte neue Methode setzt *Distant Watching* um und versucht, automatisiert den Bildinhalt zu erfassen. Damit wird im Vergleich zu bisherigen Ansätzen, die entweder nur sehr generische Informationen wie Schnittkurven (Howanitz 2015) oder Farben (Burghardt/Wolff 2016) herauslesen bzw. ganz auf manueller Annotation beruhen (Dunst/Hartel 2016), eine wesentliche Verbesserung erreicht. Ein State-of-the-Art *Regional Convolutional Neural Network* (RCNN) wird auf konkrete vorselektierte Symbole in Videos trainiert, um diese in einem großen Videokorpus automatisiert erkennen zu können. Damit wird erstmals der Bildinhalt von Videos automatisiert erfass- und quantitativ messbar. Je nach (Co-)Präsenz oder Absenz von Symbolen können Rückschlüsse auf den Inhalt des Videos gezogen werden.

Diese Ausweitung des methodischen Repertoires auf Bilder bzw. Videos ermöglicht den Kulturwissenschaften quantitative Perspektiven auf Malerei, Photographie und Film. Auch Objekte oder performative Handlungen können über Bild- bzw. Videodokumentationen einer quantitativen kulturwissenschaftlichen Analyse zugeführt werden. Diese quantitative methodische Innovation muss allerdings durch eine qualitative ergänzt werden. Die vorliegende Studie setzt sich zum Ziel, diese Innovationen anzustoßen.

MultiPath Network

Unsere Studie beruht auf einer Weiterentwicklung eines *Convolutional Neural Networks* (CNN). Ein konventionelles CNN, beispielsweise *VGG19* (Russakovsky et al. 2015), ist in der Lage, ein Eingabebild anhand eines vorher durchgeführten Trainings in genau eine vordefinierte Klasse einzuordnen. Für ideale Eingabebilder, etwa jene aus *MNIST*- (Lecun et al. 1998) oder *CIFAR100-Korpus* (Krizhevsky 2009), ist das ein praktisch gelöstes Problem. Enthält das Bild allerdings Instanzen mehrerer Klassen, produziert ein konventionelles CNN keine verwertbaren Ergebnisse. CNNs wurden deshalb zu *Regional Convolutional Neural Networks* (RCNN) weiterentwickelt, wie beispielsweise zu *MultiPath Network* (Zagoruyko et al. 2016).

RCNNs operieren zweistufig: Zuerst werden automatisch Regionen in einem Bild vorgeschlagen und intern noch verfeinert. Anschließend werden diese vorgeschlagenen Ausschnitte klassifiziert. *MultiPath* ist standardmäßig durch die generischen Bilder des *COCO*-Korpus (Lin et al. 2014) vortrainiert und erkennt alltägliche Dinge (Katzen, Flugzeuge, Speisen, usw.). Wie kleine explorative Experimente gezeigt haben, ist es notwendig, auf dem allgemeinen Training aufbauend eigene Trainingsläufe für jene visuellen Symbole zu entwickeln und durchzuführen, die uns für die konkrete kulturwissenschaftliche Fragestellung interessieren. Dies bedeutet einen hohen Aufwand an Rechenzeit und verlangt spezialisierte Hardware; dafür lassen sich RCNNs dann aber auf jegliche Arten von Symbolen oder andere visuell unterscheidbare Merkmale trainieren und können zu deren Identifizierung eingesetzt werden.

Stepan Bandera

Zentrum der Untersuchung dieses Papers ist die Rezeption des ukrainischen Nationalisten Stepan Bandera, die in sich die Ambivalenz ukrainischer Erinnerungskultur vereint und die im gegenwärtigen Ukrainekonflikt immer wieder polarisiert: Für das prorussische Lager ist er ein Faschist und Massenmörder, seine Anhänger werden als *Banderovcy* mit Faschisten gleichgestellt. Für die ukrainisch-nationalistische Seite ist Bandera ein idealisierter Held, der kompromisslos für die nationale Unabhängigkeit kämpfte. Neue Medien werden intensiv genutzt, um die von der jeweiligen Seite präferierte Sicht auf Bandera durchzusetzen. Eine erste Untersuchung zeigte, dass sich diese Instrumentalisierung durch alle größeren digitalen Medien zieht und bereits vor 2014 immanent war (Fredheim et al. 2014). Unser Paper baut auf dieser Vorarbeit auf; wir vergleichen das Youtube-Korpus vor dem Kriegsausbruch in der Ukraine mit einem heutigen Korpus, um aufzuzeigen, ob und wenn ja, wie der Ukrainekonflikt die bereits vorhandene unterschiedliche Instrumentalisierung verändert hat. Als Korpus dienen uns die jeweils 200 ersten *YouTube*-Suchresultate für die Begriffe "Stepan Bandera" und "###### #######". Dass die *YouTube*-Suche keine objektive Übersicht über den Datenbestand liefert, sondern die Ergebnisliste je nach Land, Browser und anderen Details des Suchenden anpasst, sei angemerkt, kann an dieser Stelle allerdings nicht weiter diskutiert werden.

Neben der propagandistischen Instrumentalisierung ist ebenso auf die Ebene der "post-memory" (Hirsch 2012) zu verweisen. Marianne Hirsch beschreibt mit diesem Konzept eine Auseinandersetzung mit einer traumatischen Vergangenheit, die man selber nicht erlebt hat. Dabei spielen visuelle Medien eine entscheidende Rolle, weil sie, so Hirsch, emotional aufladbarer sind als Texte. Wie diese emotionale Komponente im Rahmen des *Distant Watchings* mitbedacht werden kann, ist sowohl aus qualitativer als auch als quantitativer Sicht zu klären.

Methode

Automatische Lokalisierung und Klassifizierung erfordern eine genaue Definition der Objekte, die gefunden werden sollen. Wir haben eine Reihe von 12 typischen Symbolen festgelegt, die im Kontext der Auseinandersetzung über Bandera häufig verwendet werden, darunter Symbole des russischen oder ukrainischen Nationalismus bzw. des Faschismus.

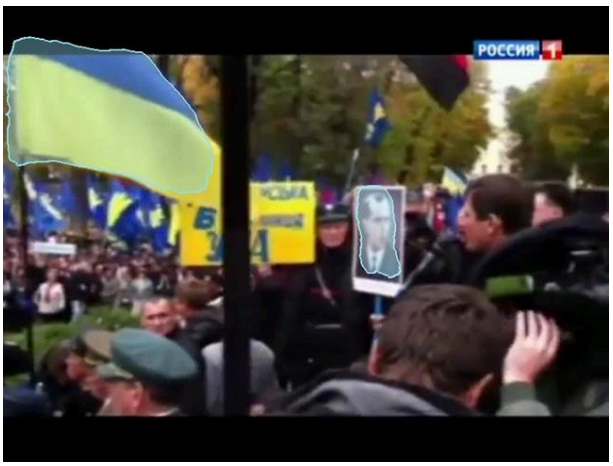
Symbole des ukrainischen Nationalismus:	Symbole des *Faschistische Symbole:*	Symbole des polnischen Nationalismus	Symbole des *russischen (sowjetischen) Nationalismus:*
---	---	---	---
ukrainisches Wappen (182)	Hitler-Bilder (95)	polnisches Wappen (38)	Hammer & Sichel (111)
Bandera-Bilder (110)	Hakenkreuz (190)	Falanga (95)	Georgsband (147)
ukrainische Flagge (168)	SS-Rune (107)		
Flagge der UPA (48)			
Swoboda-Symbol (129)			

Die Auswahl der Symbole erfolgte aus einem Close Watching einer Reihe von Beispiel-Videos zu Bandera heraus. Es handelt sich hierbei um wiederkehrende Symbole, die klar dafür genutzt wurden, um eine wertende Aussage im Bildprogramm zu platzieren. Die Symbole werden manuell anhand von Bildern aus den Beispielvideos sowie an Bildern aus dem Internet annotiert. Bisherige Tests zeigen, dass die Symbole auf mindestens 80 Bildern annotiert werden müssen, um robuste Ergebnisse erzielen zu können. Die automatische Klassifikation erlaubt es, das gesamte Korpus nach den trainierten Symbolen durchsuchen zu lassen. Außerdem gibt es Auskunft, wie viel Platz es in dem Video eingenommen hat und wie lange es sichtbar war. Mithilfe dieser Daten kann das Korpus statistisch analysiert und beispielsweise festgestellt werden, in welchem symbolischen Kontext Bandera gezeigt wird.

Das experimentelle Korpus umfasst 813 Bildern mit insgesamt 1483 Annotationen. Das ergibt im Mittel 123 annotierte Objekte pro Kategorie. Eine Annotation besteht aus Punktkoordinaten, die den Umriss des Objekts angeben und den zugehörigen Name der Klasse. Einem Bild sind zwischen 1 und 13 Annotationen zugeordnet. Im Durchschnitt sind es 1,7; der Median beträgt 1. Um Overfitting zu vermeiden, wird das Korpus, wie üblich, zufällig in Trainings- und Evaluationsdaten in einem Verhältnis 80/20 aufgeteilt.

Die Evalutationsmetrik der ersten Stufe wird mit dem numerischen Maß *Intersection over Union* (IoU) aus dem Intervall von 0 bis 1 angegeben. Je mehr dieser Wert gegen 1 geht, desto mehr stimmt die vorgeschlagene Region mit der vordefinierten Region überein.

Experimente mit den beiden Unterstufen der ersten Stufe (Lokalisierung von Objekten und deren Verfeinerung) zeigen einen durchschnittlichen IoU von 0,68 (Median 0,76). Für Symbole, die in deutlich über 80 mal in Bildern annotiert werden konnten ist der IoU mit 0,74 (Median 0,76) nochmals höher.

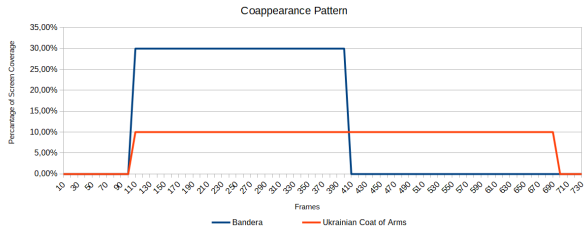


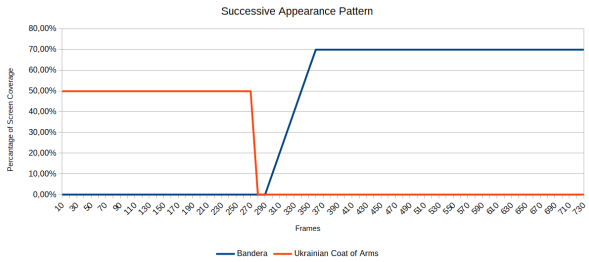
Auf den hier gezeigten Beispielbildern sind automatisch erkannte Regionen trainierter Symbole farblich hervorgehoben. Auf dem ersten Bild wurde die ukrainische Flagge sowie das Konterfei Banderas erkannt, auf dem zweiten Bild Hammer und Sichel, auf dem dritten Bild das Gesicht Adolf Hitlers. Aber nicht nur Vorhandensein und Nichtvorhandensein der Symbole ist feststellbar, auch Position und Größe können extrahiert werden.

Unsere Programme sollen es ermöglichen, Videos direkt aus dem Stream heraus in MultiPath zur Verarbeitung laden zu lassen, weil eine Speicherung der Videos auf der Festplatte augenblicklich nicht möglich ist, da dies gegen die AGB von YouTube verstoßen würde. Unsere Abwandlung von MultiPath Network erzeugt aus den Informationen des Videostreams eine Beschreibungsdatei im JSON-Format mit Informationen, welche Symbole in welchen Frames erkannt wurde und wie viel Fläche es eingenommen hatte.

Dies ermöglicht eine statistische Auswertung welche Symbole zusammen mit anderen gezeigt werden. und welche nicht. Für jedes Frame wird berechnet, wie viel Platz ein Symbol im Frame einnimmt. Diese Werte können dann für das gesamte Video ausgewertet werden, wie exemplarisch unten in zwei Bildern gezeigt wird. Solche statistischen Auswertungen ermöglichen die automatische Kontextualisierung der Bandera-Videos. Je nach Symbolen, die gleichzeitig, oder im Umfeld mit, Bandera gezeigt werden, lassen sich Aussagen treffen, ob das Video pro-russisch oder pro-ukrainisch einzuordnen ist.

Auch lässt sich auf diese Weise untersuchen, ob mit der Zeit bestimmte Symbole (zum Beispiel faschistische) in den Videos zu- oder abnahmen.





Zusammenfassung, Ausblick, Kritische Reflexion

Die Methode steht und fällt mit der Zusammenstellung der zu trainierenden Symbole. Werden wichtige Symbole beim Training außen vor gelassen, hat dies große Auswirkungen auf die interpretatorische Aussagefähigkeit. Ähnlich wie bei Texten ergeben sich auch bei visuellen Medien erst durch die Kombination von Close und Distant Watching Synergie-Effekte (Hayles 2010).

Unser Experiment konnte zeigen, dass der Ansatz, YouTube-Videos mit *MultiPath* "aus der Ferne" zu betrachten, funktioniert. Derzeit wird das Training von *MultiPath* optimiert, um bestmögliche Resultate zu generieren. Nächster Schritt ist dann die komplette Auswertung des Korpus und eine Überprüfung der Resultate durch ein Close Watching ausgewählter Videos. Die Ergebnisse dieser Verfeinerung werden mit in den Vortrag einfließen.

Bibliographie

Burghardt, M. / Wolff, C. (2016). "Digital Humanities in Bewegung. Ansätze für die computergestützte Filmanalyse" in *DHd 2016: Book of Abstracts*, 108-112.

Dunst, A. / Hartel, R. (2016). "Die Corpusanalyse multimodaler Erzählungen am Beispiel graphischer Romane" in *DHd 2016: Book of Abstracts*, 120-122.

Fredheim, R. / Howanitz, G. / Makhortykh, M. (2014). "Scraping the Monumental: Stepan Bandera through the Lens of Quantitative Memory Studies" in *Digital Icons* 12 (2014), 25-53. http://www.digitalicons.org/wp-content/uploads/issue12/files/2014/11/DI12_2_Fredheim.pdf [letzer Zugriff 11. September 2017].

Hayles, N. K. (2010): "How We Read: Close, hyper, Machine" in: *ADE Bulletin, 150*: 62-79.

Hirsch, M. (2012). "The Generation of Postmemory: Writing and Visual Culture After the Holocaust", New York: Columbia University Press.

Howanitz, G. (2015). "Jožin z Bažin – Ein Mem, aus der Distanz betrachtet" in: Simonek, Stefan / Doschek, Jolanta (eds.): *Slawische Popkultur*. Wien: PAN, 63-80.

Shifman, L. (2013): "Memes in Digital Culture". Cambridge (MA): MIT Press.

Krizhevsky, A. (2009). "Learning Multiple Layers of Features from Tiny Images" https://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf [letzter Zugriff 12. Januar 2018].

Lecun, Y. / Bottou, L. / Bengio, Y. / Haffner, P. (1998): "Gradient-based learning applied to document recognition" in: *Proceedings of the IEEE*, 86(11): 2278–2324 doi: https://doi.org/10.1109/5.726791 .

Lin, T. Y. / Maire, M. / Belongie, S. / Hays, J. / Perona, P. / Ramanan, D. / Dollár, P. / Zitnick, C. L. (2014): "Microsoft COCO: Common objects in context" in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS. pp. 740–755 doi:10.1007/978-3-319-10602-1_48. http://arxiv.org/abs/1405.0312 [letzter Zugriff 12. Januar 2018].

Russakovsky, O. / Deng, J. / Su, H. / Krause, J. / Satheesh, S. / Ma, S. / Huang, Z. (2015): "ImageNet Large Scale Visual Recognition Challenge" in: *International Journal of Computer Vision*, 115(3): 211–252 doi:10.1007/s11263-015-0816-y.

Zagoruyko, S. / Lerer, A. / Lin, T.-Y. / Pinheiro, P. O. / Gross, S. / Chintala, S. / Dollár, P. (2016): "A MultiPath Network for Object Detection". *BMVC* http://arxiv.org/abs/1604.02135 [letzter Zugriff 12. Januar 2018].