

Applying Computer Vision Systems to Historical Book Illustrations: Challenges and First Results

Yongho Kim¹, Thomas Mandl¹, Chanjong Im¹,
Sebastian Schmideler², Wiebke Helm²

¹ University of Hildesheim, Information Science, Germany

² University of Leipzig, Faculty of Education, Germany
mandl@uni-hildesheim.de

Abstract. Digital humanities still need to unlock the potential of images analysis algorithms to a large extent. Modern deep learning images processing can contribute much to quantify knowledge about visual components in books. In this study, we report on experiments carried out for historical print. The illustrations in books offer much for humanities research. Object recognition systems can identify the portfolio of objects in book illustrations. In a study with several hundreds of books, we applied systems to find illustrations and classify them. Results show that persons are shown in illustrations within fiction books with a higher frequency than in non-fiction books. We also show the classification results for an analysis of the printing technology. This expert task can still not be perfectly modeled by a CNN. A class activation map analysis can be used to analyze the performance qualitatively.

Keywords: Digital Humanities, Children Books, Deep Learning, CNN.

1 Introduction

Digital Humanities integrates automatic processing and analysis into research practices in the Humanities. Image analysis is a growing area within Digital Humanities. The analysis of books is of great interest to many disciplines. Digital historical corpora allow the automatic access to illustrations in books and their analysis in large quantities. This can lead to innovative research questions and quantitative results [1].

Historical children's and youth books have yet not often been the subject of research. Children books typically contain more images than adult books typically. As a consequence, they are of special interest for an analysis of images. In addition, they form a closed category on the one hand which contains sufficient variety on the other hand [2].

Illustrated books have played a significant role in knowledge dissemination. The declining production costs for printed images have led to a growing exposure of more and more people to rich visual resources. Research in this area could identify trends in the objects depicted.

2 State of the Art

The progress in computer vision and images analysis in the last decade was substantial. Deep learning as a new direction in machine learning now can be considered as state of the art and delivers excellent results in many domains. Deep learning refers to a departure from feature engineering. Algorithms instead find a representation space for the problem at hand by themselves. Neural networks have proven to be very successful for this task. Already early approaches intended to input compresses feature spaces into backpropagation neural networks [3] but recently, new architectures have grown more complex and successful.

Convolution Neural Networks (CNN), the recent state of the art technology is known to be very effective in automated feature detection and subsequent classification in many domains. CNNs are composed of recurring sets of two layers: a convolution layer and a pooling layer. The CNN combines pixels locally and by working through many layers, more complex features can be extracted [4].

The applications of computer vision to the digital humanities is a growing area. However, there are still not very many publications in this domain. One influential experiment in the art domain by Salah & Elgammel is dedicated to classify the painter of artistic work. Such work is highly dependent on the type of paintings in the collection [5]. An approach to identify objects within art work has also been presented. Similar to our approach, it needs to deal with the domain shift and apply current technology to historic print [6].

A recent project is focusing on research on graphic novels. Current state of the art CNNs are applied to tasks like author identification with very good success [7]. In addition, the processing is aimed at measuring the drawing style of a graphic novel in order to find similar book titles. A study of modern children books based on information available in catalogues has analyzed market structures and book formats [8].

A recent approach shows that the visual analysis of a page structure can be carried out successfully with CNNs. A system can detect elements on a page and analyze tables from heterogeneous layouts [9].

One goal for the research of images in historical children books lies within the production technologies. As a classification problem with few classes, it seems like a challenge which could be solved with current technology. The classification of production technology in the 19th century is still a hard task [10]. Comparable work is the application of deep networks and transfer learning for material classification. Cimpoi et al. conducted material classification with deeper structure and transfer learning [11].

3 Data Collections of Digitized Books

This research is exploiting two collections of books that are digitized. The first collection is the Wegehaupt corpus maintained by the Staatsbibliothek in Berlin [12]. The second data collection is based on the Hobrecker collection [13]. This collection of books is maintained by the library of the Technical University of Braunschweig.

All collections are of great interest for cultural research. They contain a rich variety of different genres of children books mainly from Germany and mainly from the 19th century: e.g. alphabetization books, picture books, biographies, natural history descriptions as well as adventure stories. Another resource used is the database *pictura paedagogica online* [14]. It contains only images, many of which are extracted from books.

4 Results and Analysis

We present some results for using a system with pre-defined classes and one with a self-trained model. All models use deep learning systems and in particular some variant of a CNN architecture.

After extracting images from the book pages, we use them as input for an object classification system pre-trained on modern photographs. In previous work, we could see that the classification results are not perfect and that they greatly differ between books. Typical performance metrics can lie between 30% and 90% for different books [15] which shows the heterogeneity of the material and the sensitivity of the systems for that.

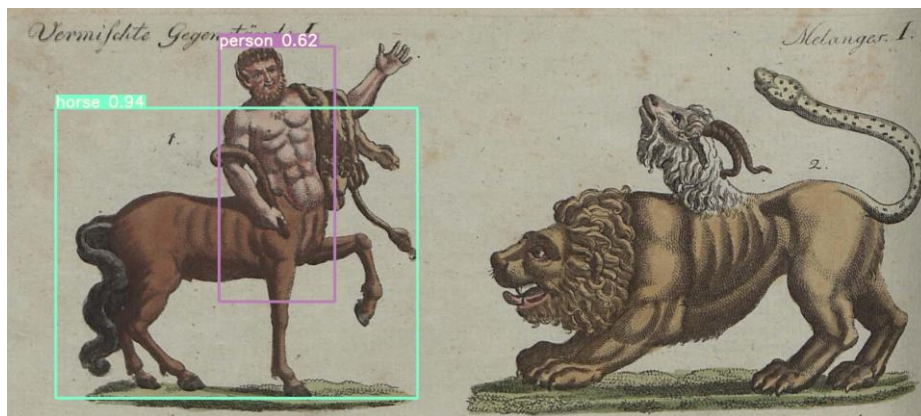


Fig. 1: Example of an illustration from the collection with object recognition results.
Bertuch (1801). Bilderbuch für Kinder.

Yolo to the images and record the recognized object types. The results of Yolo [16] have been recorded for a subset of 321 books from the Hobrecker collection. The classification is not very reliable; a detailed evaluation is ongoing. However, for a statistical analysis, it seems sufficient. We manually classified the fiction books and non-fiction books. The analysis shows that there is no difference in the overall number of illustrations for the two classes. However, the fiction books contain more images of humans and horses and thus a more limited scope of object classes than is the case in the non-fiction books. Non-fiction displaying many different objects and animals seem to cause that difference (see table 1).

Table 1. Results of Yolo on the Hobrecker collection.

<i>Median</i>	<i>Illustr. per page</i>	<i>Person per illustr.</i>	<i>Horse per illustr.</i>	<i>Chair per illustr.</i>
fiction books	0.400	1.143	0.020	0.0
non fiction books	0.526	0.737	0.042	0.0

<i>Average</i>	<i>Illustr. per page</i>	<i>Person per illustr.</i>	<i>Horse per illustr.</i>	<i>Chair per illustr.</i>
fiction books	0.448	1.446	0.103	0.072
non fiction books	0.566	1.151	0.128	0.035

It needs to be stressed that the distribution of the classes is highly skewed. The most frequent classes are humans and a few animals. These results currently do not allow a quantitative tracing of many different motifs through the century.

For the study of historical print, questions of materiality are of great importance. Issues of aesthetic design needs to be considered in relation to the techniques and printing technologies available. Printing technology like wood cut, wood engraving and lithography allowed different levels of elaboration. Finding the technique is a tedious task. It is often not stated in the meta data and it requires experts to identify it from digitized books.

Therefore, the automatic identification is a important requirement. We trained a model for distinguishing between three classes and managed to achieve only a reasonable performance. However, for a statistical analysis, this can be sufficient.

Table 2. Results 3 models for a 3-class classification for printing technology. First number reports accuracy on the training set, second number on the test set.

Model	Layers	RGB	Sym. tran.
ResNet20	20	98.3%/64.6%	99.3%/64.9%
ResNet56	56	97.7%/60.4%	99.2%/58.6%
Shallow/Parallel	3	72.5%/63.9%	74%/64.9%

To further analyze the errors and to observe how the algorithms differs from human experts, we applied a qualitative analysis with class activation mapping (CAM)[17]. They show the areas which were relevant for the system to classify the image into a class. We can see that the algorithm does often not consider content parts of the image but rather parts of the frame. Also, experts look at long lines or large areas.

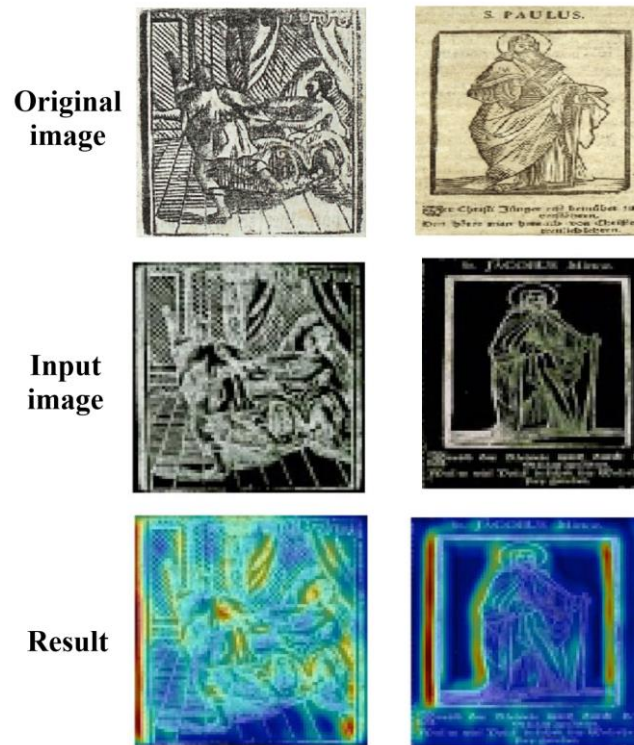


Fig. 2: Example of an CAM analysis for images from the collection.

5 Outlook

Future research needs to also address stylistic and artistic aspects of illustrations. A deeper analysis of content on a page and in particular of frequent classes (primarily pictures of humans) offer great potential for advanced analysis tools for digital humanists. We intend to develop a scene detection system which allows the study of typical scene types like family, play, schooling and nature.

Acknowledgements

We thank the Fritz Thyssen Foundation for their funding for the research project *Distant Viewing*. We thank the library of the *Technische Universität Braunschweig*, the *BBF | Bibliothek für Bildungsgeschichtliche Forschung Bibliothek Bildungsgeschichtliche Forschung, Abteilung des DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation* and the *Staatsbibliothek Berlin (Preußischer Kulturbesitz)* for providing and facilitating access to their digitized collections.

References

1. Bentkowska-Kafel, A. (2015). Debating Digital Art History. In: International Journal for Digital Art History. vol. 1 <https://doi.org/10.11588/dah.2015.1.21634>
2. Schmideler, S. (2018). Lutherbilder. Ein Streifzug durch die Illustrationsgeschichte der Kinder- und Jugendliteratur des 18. und 19. Jahrhunderts. *Die Reformation in der Kinder- und Jugendliteratur*.
3. Mandl, T. (2000). Tolerant information retrieval with backpropagation networks. *Neural Computing & Applications*, 9(4), 280-289.
4. Skansi, S. (2018). Introduction to Deep Learning. Springer.
5. Saleh, B. & Elgammal, A. (2016). Large-scale Classification of Fine-Art Paintings: Learning the Right Metric on The Right Feature. *Intl. Journal for Digital Art History*, (2).
6. Crowley, E. & Zisserman, A. (2014). The State of the Art: Object Retrieval in Paintings using Discriminative Regions. *Proc. British Machine Vision Conference*. BMVA Press.
7. Dunst, A. & Hartel, R. (2018). Auf dem Weg zur Visuellen Stilometrie: Automatische Genre- und Autorunterscheidung in graphischen Narrativen. *Kritik der digitalen Vernunft*. 5. Tagung „Digital Humanities im deutschsprachigen Raum“ <http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf>
8. Steiner, A. (2019). Conservatism in an Innovative Field: Children’s Digital Books in Sweden. *DHN 2019 Digital Humanities in the Nordic Countries 4th Conference*. ceur-ws.org/Vol-2364
9. Lehenmeier, C., Burghardt, M., & Mischka, B. (2020, August). Layout Detection and Table Recognition—Recent Challenges in Digitizing Historical Documents and Handwritten Tabular Data. In *International Conference on Theory and Practice of Digital Libraries* (pp. 229-242). Springer, Cham.
10. Im, C.; Ghauri, J.; Rothman, J. and Mandl, T. (2018). Deep Learning Approaches to Classification of Production Technology for 19th Century Books. LWDA, pp. 150-158. <http://ceur-ws.org/Vol-2191>
11. Cimpoi, M., Maji, S., Kokkinos, I. & Vedaldi, A. (2016). Deep filter banks for texture recognition, description, and segmentation. *International Journal of Computer Vision*, vol. 118 (1) pp. 65-94
12. Staatsbibliothek zu Berlin, Preußischer Kulturbesitz. Wegehaupt Digital: [https://digital-beta.staatsbibliothek-berlin.de/suche?category\[0\]=Kinder- und Jugendbücher&queryString=project%3A\"wegehauptdigital\"](https://digital-beta.staatsbibliothek-berlin.de/suche?category[0]=Kinder- und Jugendbücher&queryString=project%3A\).
13. UB TU Braunschweig, Hobrecker Kollektion Online. https://publikationsserver.tu-braunschweig.de/content/collections/childrens_books.xml
14. Jornitz, S., & Kollmann, S. (2004). Ins Bild hinein und aus dem Bild heraus. Anmerkungen zu Erfahrungen im Umgang mit einer pädagogischen Bild-Datenbank. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 9, 1-17.
15. Mitera, H.; Im, C.; Mandl, T.; Womser-Hacker, C. (2021). Objekterkennung in historischen Bilderbüchern: Eine Evaluierung des Potenzials von Computer Vision Algorithmen. In: –KinderBuch. Reihe Studien zu Kinder- und Jugendliteratur und -medien. J.B. Metzler.
16. Redmon, J.; Divvala, S.; Girshick, R. and Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788.
17. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision* pp. 618-626.