Deep Learning Approaches to Classification of Production Technology for 19th Century Books

Chanjong Im, Junaid Ghauri, John Rothman, Thomas Mandl
University of Hildesheim, Germany, uni-hildesheim.de

Context and Goals

- Digital Humanities mainly operate on text collections
- Images are an important facet in knowledge dissemination
- Cooperation Project *Distant Viewing* with the University of Leipzig
- The 19th century saw a huge growth in the production of images in books
- Children Literature is a good example and a closed genre
- The University Library of the TU Braunschweig provided a digitized collection (Hobrecker)

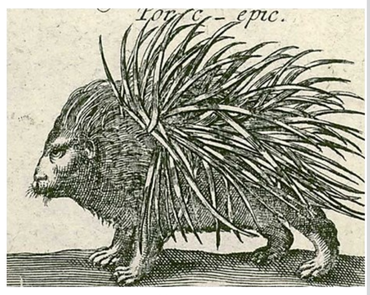
Printing Technologies

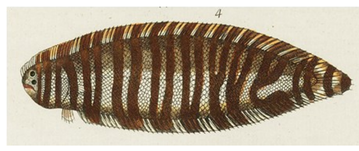
Main Technologies

- Wood Cut
- Wood Engraving
- Copper Engraving
- Lithography

Hard to distinguish for humans

Effects on style and content





Classification Experiments with CNNs

https://digital.staatsbibliothek-berlin.de/werkansicht?PPN=PPN757444229

Name	Accuracy
Big-Filters	63%
Inception Model with Neural Networks	58%
Inception Model with linear SVM	47%
Inception Model with non-linear SVM	52%
Small-Filters-Less Pooling	61%
Small-Filters-Balanced Pooling	48%

Research Challenges

- Do methods from texture recognition work?
- Are methods and pre-trained models derived from realistic photographs adequate for 19th century illustrations?
- Are full re-sized images or cropped part with full detail better for the task?

Future Research

- Similarity Analysis to research patterns of re-use of images
- Object Identification to research knowledge dissemination about innovative technology or exotic animals
- Stylometric analysis (Are there typical patterns for publishing houses?)

Partner: Universität Leipzig

Funding: Fritz Thyssen Stiftung für Wissenschaftsförderung

For further information contact: mandl@ or imchan@uni-hildesheim.de
https://www.uni-hildesheim.de/fb3/institute/iwist/forschung/forschungsprojekte/aktuelle-projekte/distant-viewing/