Multimedia Tools and Applications (2022) 81:5867–5888
https://doi.org/10.1007/s11042-021-11754-7

Deep learning for historical books: classification of printing technology for digitized images

Chanjong Im[1] · Yongho Kim[1] · Thomas Mandl[1]

Received: 7 June 2021 / Revised: 20 September 2021 / Accepted: 22 November 2021
Published online: 31 December 2021

Abstract
Printing technology has evolved through the past centuries due to technological progress. Within Digital Humanities, images are playing a more prominent role in research. For mass analysis of digitized historical images, bias can be introduced in various ways. One of them is the printing technology originally used. The classification of images to their printing technology e.g. woodcut, copper engraving, or lithography requires highly skilled experts. We have developed a deep learning classification system that achieves very good results. This paper explains the challenges of digitized collections for this task. To overcome them and to achieve good performance, shallow networks and appropriate sampling strategies needed to be combined. We also show how class activation maps (CAM) can be used to analyze the results.

Keywords Printing type classification · Historical image processing · Shallow CNN · Deep learning in digital humanities

1 Introduction

Digital Humanities research is focusing on enriching scholarship in Humanities and Cultural studies by employing digital methods for collecting, preserving, and analyzing artifacts. The paradigm Distant Reading [34] includes a mass analysis of literary and other text and has proven to be especially productive. Text Mining by software tools supports the analysis of written texts and is at the same time questioning the limits of the appropriateness of Artificial Intelligence as a method for supporting knowledge creation processes for humanists.

✉ Thomas Mandl
mandl@uni-hildesheim.de

Chanjong Im
imchan@uni-hildesheim.de

Yongho Kim
kimy@uni-hildesheim.de

[1] Information Science, University of Hildesheim, Hildesheim, Germany

Since the Iconic Turn, research with images and visual material has itself established within the Humanities beyond the classic image sciences like art history. For Digital Humanities, the development of appropriate tools and methods for Distant Viewing, which stands for the automatic analysis of large amounts of objects and visual data (also considering architecture and movies) with AI algorithms is still an emerging research field. Automatic processing of large amounts of image data is still limited in the Digital Humanities. On the other hand, there are large collections of digitalized books available which can support research in cultural studies.

The printing technology for books has changed significantly over time. This has been the case in particular during the nineteenth century [35]. The rapid technological advancement during this time has greatly impacted product production. On one hand, more advanced technology allowed the inclusion of more images in the books. Also, these became much more elaborated and fine-grained compared to the techniques used in the early nineteenth century. While woodcut printing did often not work much on the background, more advanced printing devices changed the style. On the other hand, harder materials were able to be used and the utilization of much less complicated technologies enabled mass production. In particular, lithography allowed mass production which led to a new book market.

This shows that the study of artistic tradition requires the inclusion and consideration of printing technology. Otherwise, one might find relations or trends which are mainly related to the technology used. It is a desideratum in Digital Humanities and Art History to have a reliable classifier available for all printing technologies. However, it is a great challenge to identify printing techniques for illustrations from historical books. Even for domain experts, this task is far from trivial and often requires a detailed inspection. The information about the printing technology used is often not available in libraries and archives. The user cannot access how the book was printed in the meta data even if the book has been digitized and its data is openly available.

Therefore, we developed a system for solving this task with modern computer vision systems.

The paper and our research approach are structured as follows. After an overview of the task and related research, we present the data collection which was used. After that, data preprocessing methods applied to historic printed images are discussed. Subsequently, sampling strategies are presented. For the classification task, deep models are applied and optimized. The results are presented subsequently. Finally, a visualization technique is employed to analyze the outcomes for some examples of images. This helps the domain experts to observe the regions which were highly important for the decision process of the neural network.

This research represents the following contributions:

- To our best knowledge, this is the first approach to successfully tackle the challenge of printing technology for historical images. We claim that the task needs to utilize micro-level features which makes is different from the model development trend in Computer Vision.
- We present optimal settings to resolve the task which includes the proposed shallow architecture (SPCNN), pre-processing techniques and an appropriate sampling strategy.
- We show that shallow networks deliver superior performance for this task when compared to deeper networks. This shows that the decisive patterns the printing technology are more at the micro-level.

- A visualization based on CAM (class activation map) shows that regions within images for decision making are often outside the content-rich areas. This seems to be similar to human decision making for this task.

1.1 Printing Technology and its Development

The invention of Gutenberg around the year 1450 and the invention of offset printing mark major disruptive moments in the history of technology for knowledge distribution [2]. However, also in between the printing press underwent technological advancement. We briefly sketch some relevant milestones for printing images.

The woodcut is the oldest image printing technique and has been used already before the printing press was invented. It represents a relief technique in which the image is drawn on wood and the parts which should not receive ink are carved out. The elevated parts receive color and allow the transfer of the ink to the image on paper. The technique which is also called xylography relied on cheap material which could easily be carved. However, the woodcut was rather soft and the print quickly lost quality after many copies.

Wood engraving represents an *intaglio* technique in which the ink flows into the carved or deep parts of a material. It became highly popular in the arts after 1500. Another intaglio technique is the copper engraving which used the relatively soft material copper for carving out. It has been the technique of choice for most book illustrations until the beginning of the nineteenth century [51].

The arrival of lithography revolutionized printing in the nineteenth century. It enabled print with stones which were much harder than the previously used materials. Thus, lithography enabled mass production since the quality of the print did not deteriorate after many prints [1]. Lithography is based on the mutual repulsion of water and oil. After painting an image with oil-based colors, the acidified liquid is applied and penetrated the pores of the stone or later metal. This layer does not allow the original image to accept the printing ink during the final printing process.

Many collections of historic books have meanwhile been digitized in cultural institutions, However, metadata about the printing technique in library catalogs are typically not reliable. Often different technologies were used within one book. The cover and iconic parts in the first pages were printed more elaborately than images within the book. The identification of the correct technology is an expert task that requires experience and is hard to carry out with only the digital copies available. Therefore, support systems that label images correctly, should be developed in order to facilitate the research by scholars in the digital humanities.

2 State of the Art

In the last years, considerable progress has been made in image processing, especially through approaches of so-called *Deep Learning*. These data-driven methods have performed well for many tasks and have often replaced traditional image processing based e.g. on color and shape analysis [7]. These algorithms learn aspects of the pictures that need to be analyzed for the best results. Such feature learning is typical for deep learning. An importantsystem is the Convolutional Neural Network (CNN [26]) which combines many simple neurons as processors into elaborated architectures of layers. A basic CNN is composed of recurring sets of layers which include convolution, pooling layers, and non-linear activation functions. A CNN first combines pixels locally and by working through many layers, more complex features can be extracted [43]. Based on the features, diverse classification tasks can be learned by these neural networks for image processing.

The remainder of this section is structured as follows. First, issues of deep learning classification which are relevant for our application are discussed. Then, prominent examples of image classification within Digital Humanities are presented. The last section presents some work which leads into the direction of classifying printing technology. This includes one previous publication and some similar work in the domain of low level feature identification using deep learning models.

2.1 Deep Models for Image Classification

Since the initial CNN models (e.g. [26]), systems have increased in complexity and in particular in size. The models with few convolutional layers and heterogeneous filters, often in parallel structures have shown significant improvements over the traditional image processing methods (AlexNet [24], VGG [41], GoogLeNet [46]). These early models have seen further enhancements by adopting deeper architectures containing 100 layers or more [14]. However, the deeper models typically suffered from the vanishing gradient problem. This was elevated by ResNet [14] and DenseNet [17] which have introduced residual and dense connections, respectively. The application of these types of connections has shown improvements in the classification performance. Nevertheless, the advantages of adding more layers and making the model deeper require a better understanding of the systems. Although using deeper structures and bigger models often shows better performance, the accuracy gains are often not substantial [48]. The effect is claimed to be minimal, while the details and parameters of models seem to have more influence on the robustness of models [45].

Despite the success of deep architectures, numerous shallow networks have been proposed over the last years. The general motivation of using shallow networks is to reduce the time consumed during model training while keeping the accuracy similar to that of deep architectures. Also, they are used for avoiding the overfitting problem which particularly occurs when utilizing a limited amount of data [29]. A study [12] reports the partial success of shallow network applications on few different image datasets. The model architectures that contain only one or two convolutional layers were utilized. The relative success was shown on SAT, Brazilian Coffee, and MNIST datasets but for CIFAR and UC Merced Land Use the results were comparatively less. The shown success was assumed to be from the common intraclass low-level features such as color or shape. This study is further extended by Lei et al. [28] who showed a successful application of shallow networks on MNIST by comparing them to deep architectures. Similar effects were shown for the F-MNIST dataset [13]. Li et al. [29] report the successful application of shallow networks for six apple types classification task. Lower-level features such as color, shape contour, and surface texture were considered to be more discriminative than the higher-level semantic features for this task.

The study by Hossain et al. [16] highlights the inefficiency of deep architectures and proposed a greedy algorithm to find the optimal width and number of layers in the CNN. These studies commonly claim that using an excessive number of convolutional layers is not efficient. Instead they argue that finding the optimal number of layers and number of filters leads to better efficiency and often performance.

Research by Tan et al. [48] reports that the classification performance is dependent on all types of CNN model scaling configurations. These relate to finding optimal depth, width, and image resolution coefficient values. The study shows the efficiency and performance improvements when all the values are correctly adjusted (i.e. compound scaling). This is shown using the eight variants of the proposed EfficientNet architecture which is constructed by the AutoML. The variants contain scaled model architectures which are referred to as EfficientNetB0 to EfficientNetB7, where B0 contains the smallest and B7 the largest number of parameters. Using these variants, they show similar performance achievements to the other models in their experimental group using the significantly smaller amount of parameters. Also, they were able to obtain an improved domain transfer performance. Furthermore, they showed that their compound scaling strategy works well when applied to other former mainstream model architectures such as ResNet [14], Inception-V4 [47], etc. However, these experiments were all performed using datasets which contains photographs (e.g. ImageNet [8]). Even the reported domain transfer results are limited to applying the ImageNet-based pretrained models to other photographs datasets. The coefficient values need careful adjustments when applied to tasks in the Digital Humanities.

The composition of the convolutional layers and the choice of various filters are known to play a vital role in extracting relevant features that critically contribute to the model's performance. The complex features including both detailed and high-level abstractions are learned by these choices. The nature of convolutional layers induces the increased receptive fields which in turn produce varied levels of feature abstractions. These levels and the type of features vary depending on the depth, width, input resolution, and selection of hyperparameters [48]. Such characteristics are well utilized in some applications e.g. for style transfer [11] and domain adaptation [53]. However, it remains a challenge to foresee and fully anticipate the outcome of the features derived from the deep model architectures. These limitations lead to various challenges especially in the attempts to solve new types of problems as in the work presented here. For example, the effect of using pretrained models for printing technology classification has not been studied nor has the effect of different compositions and settings as in [48].

More recent research [10, 37, 50] makes use of transformers [52] which originate from Natural Language Processing (NLP). Although they are claimed to exhibit better robustness and generalization than the CNNs, they typically require massive amounts of data for training [10] which makes the application infeasible in the present work.

2.2 Applications of image processing in Digital Humanities

Most of the large-scale research in image processing is currently being carried out for photographs. Such collections differ greatly from the non-realistic drawings and illustrations which can often be found in arts and historic Digital Humanities (DH) projects. It is necessary to investigate how methods like CNNs can be optimized for such tasks.

The work on image analysis in DH can be categorized in the following classes:

- Visualization approaches
- Detailed analysis of small sets of images
- Search systems, often based on similarity
- Classification systems for large amounts of images
- Analysis systems for identifying trends or other patterns

In this short overview, we will focus on classification tasks and tasks which extract basic features of books or images. The analysis of the page is a typical task. One line of research is focused on identifying the positions of text and image blocks within a page. The HBA data challenge for old books intends to improve algorithms for this task at the pixel level [32]. The best system has achieved an F-value of 50%. No deep learning models were applied. Challenges are the heterogeneous formats and genres. A study for the layout analysis of historical newspapers has been conducted which achieved very good results e.g. [27]. One study explores visual trends in newspapers and models them as a multimodal construct consisting of text and images. The similarity of images is also explored [54]. Further basic operations of visual analysis of book data include OCR e.g. [36].

Since classification with clear classes is a typical task for computer science, optimization of deep learning algorithms has been explored for this topic [55] and even benchmarks have been developed [44]. For example, Sandoval [39] has applied a 2-stage learning process.

The identification of objects within images or illustrations can be seen as a subset of this task [6]. A thorough analysis of object detection systems for a collection very similar to the one processed in our work, the differences in the recognition rate were extremely large [33]. This shows the influence of the style and state of the book. Humans are most often recognized in our collection [21].

A detailed analysis has many facets in the DH and explores different features. An analysis is carried out of the visual features of the furniture and its relation to metadata in an ontology [9]. Gesture and posture analysis within figures in art has been explored by [19]. However, one needs to consider that concepts in DH are not always clearly defined but fuzzy. Classification approaches are aiming at such high-level concepts like an art period [38] or aesthetic concepts [4].

2.3 Printing Technology Recognition and Related tasks

Previously, there are no experiments large-scale analyses of printing technology in historic print. In one study, an experiment for the two classes *Lithography* and *Woodcut* was conducted. The Inception Network architecture was applied with different filter sizes. No satisfying performance was achieved and only 63% of images in the balanced dataset were correctly classified [18].

This shows that is a need for better understanding and solving the task of identifying the printing technology.

The identification of modern printers based on images has been the subject of study. Traditional processing is based e.g. on the frequency domain. The task was to classify photocopiers, ink printers, and laser printers. Results show that more than 90% of the images are correctly identified [40].

For a similar task of classifying several printers based on the image on the paper, large data sets are available. With CNNs, a classification accuracy of 98% can be reached [20].

The work that seems to be most similar to the classification of historical printing technologies is research in the area of texture recognition and analysis of fraud detection.

Some authors manage to classify material from visual data. The materials are different types of plastic. The algorithm learns the typical texture structure of each material based on magnified data. For a material classification task, narrow structures for a CNN were used. There has also been work towards applying deep networks and transfer learning for material classification. Cimpoi and colleagues conducted material classification with deeper structure and transfer learning [5]. Some authors achieved very successful results showing an accuracy of 99 to 99.9 percent with a specially constructed dataset i.e. CUReT. Other work on fraud detection has been published before the dissemination of deep learning models [25].

2.4 Dataset of Digitized Book Images

Libraries have been digitizing books and images in many projects. However, the support for mass analysis is often limited. Research infrastructures like DARIAH contain a variety of methods of digital text analysis. However, so far these research environments do not yet reflect the growing importance of visual information. Standards like the Open Archives Initiative (OAI) have led to tools for mass download of images and digitized books. However, there are still barriers to studies of printing technology.

For an analysis of the printing technology, it is necessary to have access to ground truth. Not for all collections, this metadata is available [15]. If it has been assigned, it is often not given in a consistent way, in a consistent field or it is not exported for mass download. Finding experts for labeling is hard and expensive. We considered two collections for our experiments which are well curated and to which we got access.

A big portion of the data is provided by Pictura Paedagogica Online (PPO) of the BBF in Berlin which contains a total of over 70,000 images mainly from the nineteenth century. The goal of the PPO collection is to provide access to images related to education [23]. The entire collection includes not only book illustrations but also postcards and photographs on various subjects. (opac.bbf.dipf.de/virtuellesbildarchiv).

The second data set is retrieved from an open illustration catalog called ‘Old book illustrations’. It contains nearly 4,000 illustrations from the eighteenth and nineteenth centuries. It is important to note that these illustrations are not provided by a single organization or an institute, rather they are gathered from different sources and published under this label (www.oldbookillustrations.com).

Image collections from digitized books are very heterogeneous which poses challenges for image processing. Due to funding issues, digitalization is not a fully planned process but within libraries worldwide it takes place over decades and different external companies work on it. The technology used for digitalization improves over the years. As a result, many different levels of quality and resolution are used. Moreover, the quality of paper that is scanned is not in identical conditions across books. Also, the original datasets contain much background noise which causes problems for some algorithms.

3 Method and Processing

This central section shows how deep learning systems were adapted to the task of identifying printing technology. After prior experiments with traditional image processing methods like frequency domain analysis, deep learning has shown to the most promising technology for the task. To perform a fully supervised classification using deep models, instance pairs of both illustration and its respective annotation are necessary. Furthermore, an adequate number of instances belonging to each class are required. However, not all images from the two databases contain information on printing techniques.

As mentioned it is often difficult even for domain experts in a historic print to correctly identify the technology. Also, the labels are very heterogeneous with some having a very small number of instances. Hence, only the data that contained a sufficient number of annotations and fulfills the minimum number of instances were selected and included in *subset A*. This subset contains three classes which are *Woodcut*, *Wood engraving,* and *Copper engraving*. The general statistics of *subset A* are shown in Fig. 1. Overall, the dataset contains 7,578 images.

3.1 Outlier Treatment

Outliers appearing in a dataset can greatly decrease the generalization capabilities of machine learning models especially when the data size is limited. They need to be processed to ease the convergence during the training process. Some outliers are removed from *subset A* and others are processed by applying noise removal. In our case, mainly three types of outliers are considered.

One type of outlier comes from the unique characteristics of the historical documents. Some illustrations do not contain enough information typically for printing type classification. The differences between the techniques are anticipated to be found in regions where the ink is painted. However, illustrations with keywords such as geometry, mathematics, etc. often contain small printed areas that cover less than 10% of the entire image size. The majority of these image pixels are filled with background noise rather than relevant parts that could be used during training. These are removed from *subset A*.

In addition to the noises, the illustrations from the same book tend to show stylistic and aesthetical similarities but they are often very different from those in other books. These resemblances and dissimilarities can be observed in low-level patterns. Those from the same book share similar noise distribution while those among the different books show very different patterns. An example is shown in Fig. 2. The four illustrations on the left denoted with `ad00289_02' and the other four placed on the right by `ad00341' are included in the same book. The images on the left commonly contain a yellowish background whereas the other four are in white. Such forms of similarities and differences

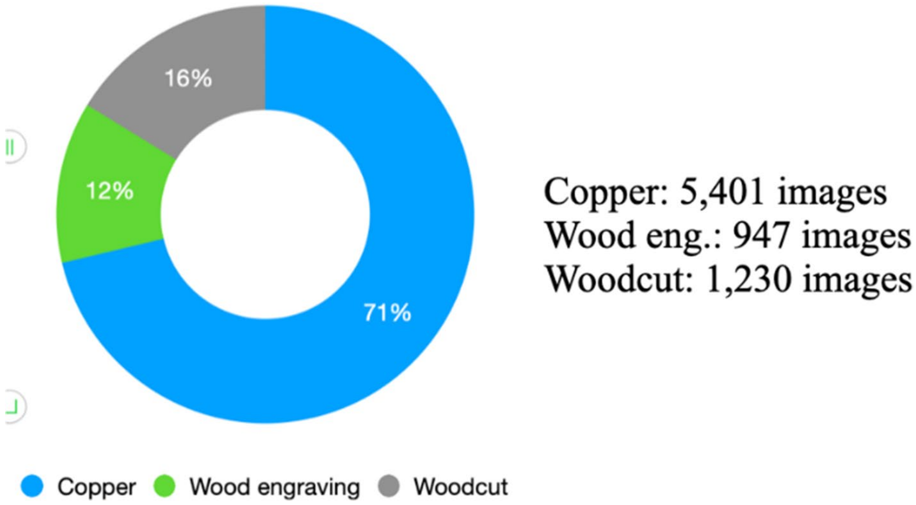
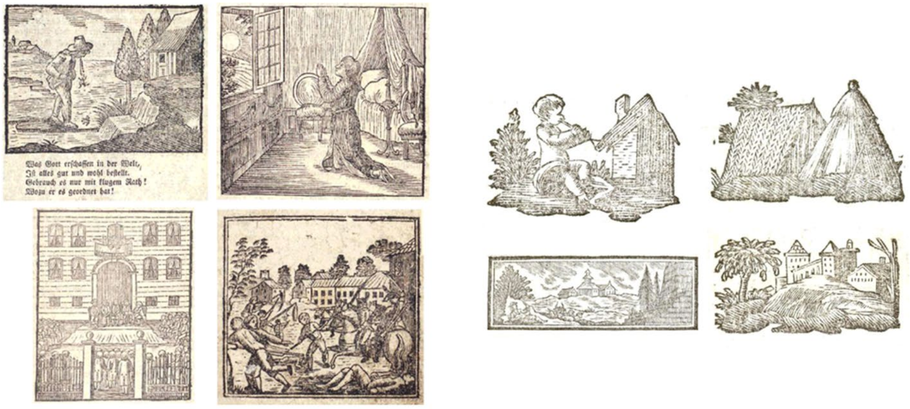


Fig. 1 Statistics of Subset A


Was Gott erſchaffen in der Welt,
Iſt alles gut und wohl beſtellt.
Gebrauch es nur mit klugem Rath!
Wozu er es geordnet hat!

ad00289_02

ad00341

Fig. 2 Image samples of two different books in the woodcut class

greatly influence the training process. The noise removal strategy mentioned in Sect. 3.3 is applied to remedy the tendencies to learn the book-specific background noises.

As our data collections contain highly diverse genres of books, some non-fiction books, for instance, contain many more images than other types of books. The books which contain a large number of illustrations can make the model generalize on the book-specific rather than relevant features for classification. A uniform amount of images from each book need to be used to avoid such biased training. Hence, two training subsets are constructed.

The first training subset is data A (D_a).

This set is constructed by randomly selecting up to ten images from each book. The intention is to reduce the data imbalance as the number of images contained in some books greatly differs from the others. The second dataset is data B (D_b) which is formed by randomly selecting up to five images from each book. The collected amount of data is relatively smaller than D_a. The main intention for extracting a reduced number of images from each book is to limit the effect of book features and closely balance the data on the book level.

The test data is formed based on different criteria. As it is meant to be used for model performance evaluation, the image uniformity for each book is disregarded. Instead, balancing the class instances is considered more important. Thus, the balanced data which contains 50 instances for each class is constructed. The same test set is used for evaluating both training sets. General statistics of D_a, D_b, and the test data are shown in Fig. 3.

3.2 Dealing with imbalanced classes

The data imbalance [42] is a typically known problem that limits the model's ability to generalize. It degrades the training ability of the models by overtraining a specific class. However, the random selection performed on both D_a and D_b incurs the class imbalanced datasets. This is remedied by adopting sampling strategies. The present work experiments the effect of the two well-known sampling methods which are *undersampling* [42] and *weighted random sampling* (https://pytorch.org/docs/stable/data.html#

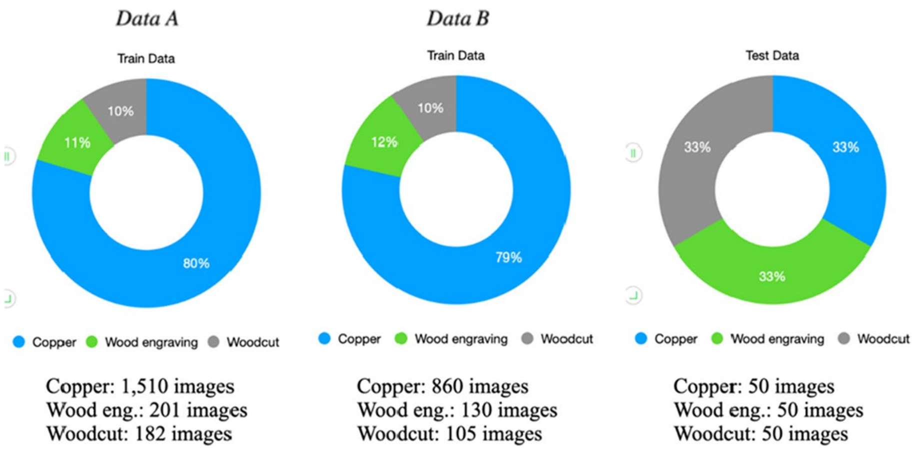


Fig. 3 Statistics of *data set A* and *data set B*

torch.utils.data.WeightedRandomSampler). Furthermore, the *epoch-wise undersampling* method is proposed and compared with the others.

The *undersampling* (US) refers to a method that attempts to establish a balanced dataset by uniformly setting the minimum amount of data instances for each class. The minimum is determined by the number of images of the smallest class. This method results in removing a substantial amount of data as the instances of the other classes are forced to be discarded. The *weighted random sampling* (WRS) involves a random image selection during the model training. All batches become balanced by selecting an even number of images per class for every batch. Similar to *undersampling*, it some training instances might never be selected in this probabilistic process based on the configured weights.

Epoch-wise undersampling (EUS) is our proposed sampling method that intends to utilize the entire training data. The process during EUS is shown in Fig. 4 and is described in the following:

1. Each class in the data is divided into subsets where each takes up as much as the number of instances in the smallest class. For instance, the data containing 10 wood engravings, 30 woodcuts, and 50 copper engravings would produce a total of 9 subsets. The subset of the smallest class, i.e. wood engraving, is formed using all instances. A total of 3 subsets is created for the woodcut class where each subset takes up 10 instances. Likewise, a total of 5 subsets is created for the copper class.

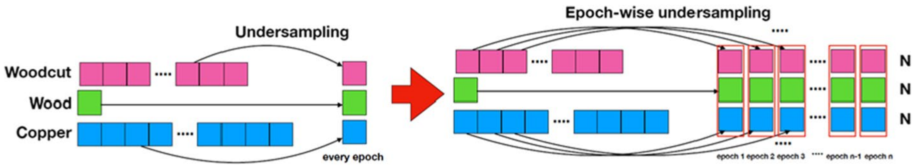


Fig. 4 Sampling strategies: *undersampling* and *epoch-wise undersampling*

2. The least common multiple L is computed using the number of subsets of all classes. In the same example L would be equivalent to 15, i.e. L = *15* (1 wood engravings × 3 woodcuts × 5 copper engravings).
3. The balanced training set is constructed which contains a total of L training instance sets. This is achieved by selecting one subset per class to form one training instance set over L iterations. Each training instance contains three subsets that are extracted from each class. Using the same example, the same subset will be selected 15 times for wood engravings. Similarly, each subset in woodcuts is selected but this is repeated five times. Likewise, each subset is selected and repeated three times for copper engravings.
4. The dataset containing L number of training instance sets is shuffled.
5. In each epoch of the training process, one training instance set is selected and used for training.
6. Steps 4 and 5 are repeated until the desired number of epochs is reached.

This procedure described that ensures the utilization of the entire training data without losing any image. This is the main difference between the *epoch-wise undersampling* and other mentioned methods which select images based on a probabilistic basis Fig. 4.

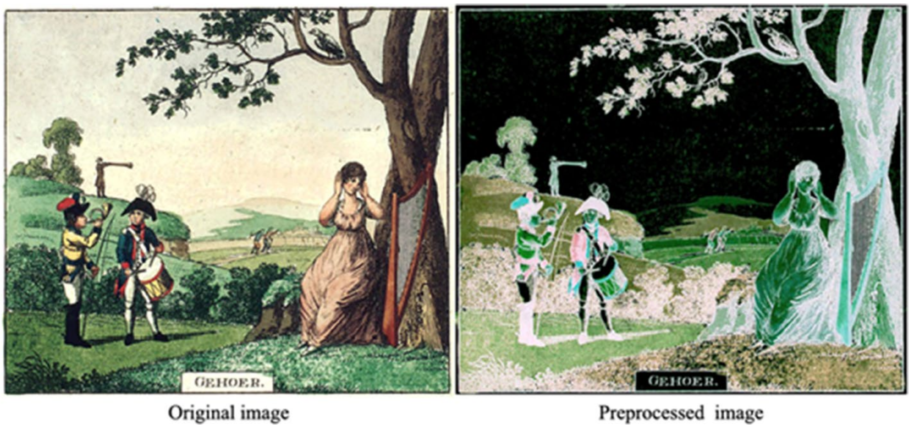
3.3 Preprocessing Methods

It is much more challenging to classify the printing technology from the digitized than from physical material. The human experts often require the unique tangible marks made by the pressing machines to examine. The visual inspection often needs very detailed and close observation. Specific and discriminative features valuable for distinguishing between the techniques are known to lie in regions with small objects of an image specifically where ink is imprinted.. The width, shape, and patterns of the lines are considered as the key distinctive features. To make the shallow network focus only on these feature types, we claim that it is critical to remove and clean the background noises that are typically seen in digitized historical documents.

The typical noise which is aimed for removal is regarded to be caused by the deteriorated paper quality of the historical books. Such type of noises tends to be present on a low level and they are assumed to affect the classification performances. Thus, the simple preprocessing technique to remove such noise type is applied and the effect is analyzed through the experiments. Methods such as Canny and Non-local Means Denoising [3] are not considered because they are known to transform pixel values. Instead, a particular and straightforward denoising method is applied (see Figs. 5 and 6) which intends to preserve details. In particular, the lamination of ink and line thickness which are considered important clues for printing technology identification can be preserved better < .

The applied noise removal preprocessing can be considered as a normalization technique and is described in the following:

- Values of an original image are transformed using the symmetric transformation function $T(x)$. The transformation is based on the condition $R(x)$ where the values lower than threshold α are replaced by 0.
- The threshold α is chosen based on the mean pixel value of the image. A higher threshold is applied if the image contains higher pixel values. In particular, if the mean value is greater than 128, $\alpha = 80$ otherwise $\alpha = 60$.



Original image Preprocessed image

Fig. 5 An example for a transformation using the proposed noise removal method

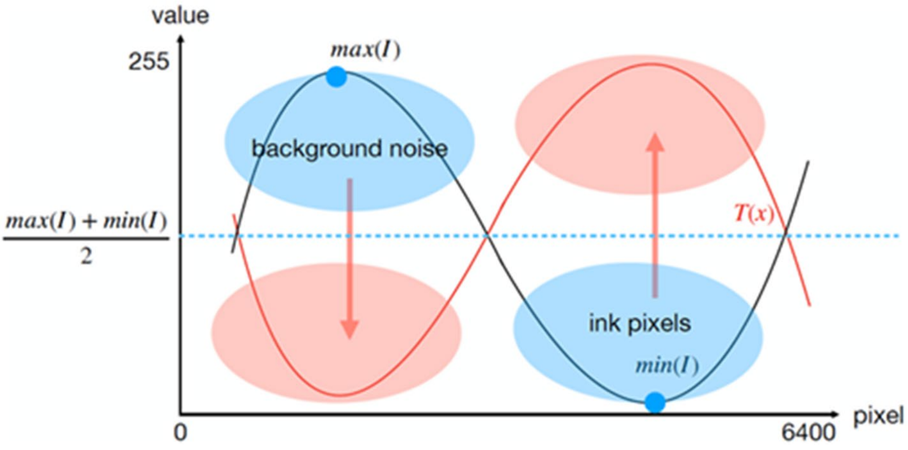


Fig. 6 Preprocessing formulas applied for an example image

To be specific, we apply a symmetric transformation to each image. It is defined as $max(I) + min(I) - x, \forall x \in I$ (I: each image) which converts high values to low values using the average of the maximum and minimum values of each image (see Fig. 6). Some transformed values are then eliminated if they are lower than the threshold α. To maximize the removal effect while trying to retain the important clues for classification, the value of α is determined based on the number of lines and ink marks present in the image. If the majority of image regions are covered with ink which will be represented with the high mean pixel value, the greater threshold of $\alpha = 80$ is applied. This is the case when the represented mean value is greater than 128. If the majority of image regions are empty, i.e. the mean value lower than 128, $\alpha = 60$ is applied.

The experiments to examine the effect of preprocessing application will be conducted using two datasets. The first data is referred to as D1 which is the image set with the

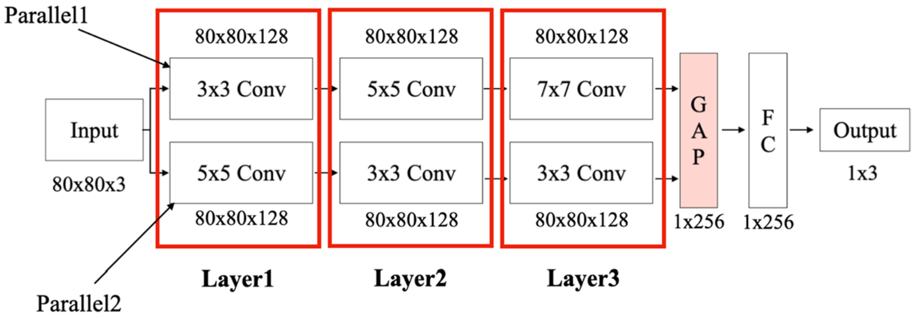


Fig. 7 Shallow and Parallel CNN architecture (SPCNN)

original RGB. The second data set is referred to as D2. It contains the images from D1 after the preprocessing method described.

3.4 Proposed model architecture

The present work intends to show the effect of using shallow CNN architecture for printing techniques classification. The designed model architecture is a Shallow and Parallel CNN (SPCNN). The architecture is shown in Fig. 7. It consists of two shallow networks each containing three convolutional layers, global average pooling (GAP), and two fully connected layers (FC).

The number of convolutional layers is chosen after a series of empirical research. The intuition behind having the network shallow is to make the model focus on narrow regions with small receptive fields. This is similar to what has been reported by Li et al. [29]. The early layers of the CNNs typically are known to pick up the simple patterns on a very limited area [31]. This is to mimic the human experts who use specific regions to differentiate the techniques.

The parallel part of the model is designed to capture the fine-grained features insusceptible to the various scan resolution. As opposed to GoogLeNet [46] and Efficient [48], (1×1) convolutional layers and constant filter size are not being utilized. Instead, various filter sizes are used which are set empirically. Note that the input size is set to 80×80. Although bigger image size and resolution are known to affect the performance, it is found to be insignificant for this particular classification from the experiments using different input sizes ranging from 30×30 to 300×300. The width in all the convolutional layers is constantly set to 128. We apply the padding method without any pooling layers. The filter size in each convolutional layer is selected based on empirical search.

3.5 Training configurations

All models including the proposed SPCNN model and other models used for comparative analyses (see Sect. 4) are trained from scratch using the Pytorch framework. The proposed model is trained using the cross-entropy loss function and Adam optimizer [22] with a learning rate of 0.0001. The mini-batch mode is used to decrease the computation overload. The size is set to 30. The other models were trained using the same configurations. However, the number of epochs was determined based on the optimal convergence level of

Table 1 Model performances using D$_a$

Sampling	US		EUS		WRS	
Model/Input	D1	D2	D1	D2	D1	D2
AlexNet	60.7%	63.3%	65.3%	65.3%	64.0%	63.3%
ResNet18	61.3%	60.7%	65.3%	62.7%	64.7%	66.7%
ResNet34	64.0%	60.7%	63.3%	66.0%	65.3%	66.7%
EfficientNetB0	78.0%	76.7%	76.0%	78.0%	76.7%	76.0%
EfficientNetB3	77.3%	78.0%	76.7%	79.3%	76.7%	76.7%
EfficientNetB5	73.3%	75.3%	75.3%	75.3%	78.7%	76.7%
SPCNN	**80.7%**	79.3%	78.7%	**82.0%**	**82.0%**	78.7%

US: undersampling, EUS: epoch-wise undersampling, WRS: weighted random sampling

D1: data without preprocessing D2: preprocessed image data

each model architecture. ResNet and EfficientNet were trained for 50 epochs. AlexNet and SPCNN were trained for 300 epochs.

To prevent overfitting or underfitting, several data augmentation methods were applied. These include random horizontal flip, random rotation, and color transformation which are available in the Pytorch framework.

4 Results

This section presents the printing type classification results using SPCNN. The results are compared with other model architectures to show the benefit of using the shallow network architecture. Further analyses show the effect of using different sampling methods, preprocessing, and data size. Finally, the regions of images which the model focused on for deciding the class are visualized.

4.1 Architecture comparison

Various model architectures were compared to the performance of SPCNN. These include AlexNet, ResNet 18, ResNet 34, EfficientNet B0, B3, and B5. AlexNet is one of the early models which is rather shallow compared to the other model architectures. It contains 5 convolutional layers and three FC layers. ResNet18 and ResNet34 are deeper architectures containing 18 and 34 convolutional layers, respectively. EfficientNet variants comprise different numbers of layers. The base model B0 is the smallest one and contains 17 layers. B3 and B5 are the scaled models from B0 using the optimal scaling values found by AutoML.

The experiment results using D_a are presented in Table 1. The rows represent the performances of each model type. The columns show the results on a twofold basis. One is the sampling method type and another is whether preprocessing mentioned in Sect. 3.3 is applied (D2) or not (D1). Note that the same test data mentioned in Sect. 3.2 is used for all experiments.

The best overall performance is achieved by SPCNN regardless of the type of sampling method or preprocessing. This indicates that this shallow architecture is much more suitable and robust for printing type classification compared to the other deeper architectures. It also reveals that the relevant features are captured on lower levels. Furthermore,

Table 2 SPCNN performance comparison based on sampling methods and the existence of data preprocessing

Sampling/Input	RS	US	EUS	WRS
D1	53.3%	80.7%	78.7%	**82.0%**
D2	42.7%	79.3%	**82.0%**	78.7%

RS: random, US: under, EUS: epoch-wise under, WRS: weighted random sampling

the results indicate that depth is not only the factor that contributes to better performance. The decreased performance is often observed for ResNet34 when compared to AlexNet or ResNet18. Also, variants of EfficientNet which considered width, depth, and image resolution for its structural design perform significantly better than AlexNet, ResNet18, and ResNet34. In this context, we argue that a parallel network equipped with filters of variable size is an adequate model architecture for our task and that it captures discriminative features that are insusceptible to the resolution and image size.

The performances of the models seem to be greatly affected by the training data types used and the sampling methods applied. However, no common tendencies are identified. A few model architectures present better performance when data without preprocessing (D1) is used and the US or EUS is applied. However, the same architecture works better when D2 is used with WRS. For example, ResNet18 shows higher accuracy results when US and EUS are applied to D1 but reversed phenomena are observed when WRS is applied. Such results are likely to be a result of the interrelation of data and model size which is closely related to overfitting, underfitting, and model complexity. Also, it could be caused by intensive preprocessing which may have removed or altered the discriminative features from the images.

4.2 Sampling methods comparison

The strength of using well planned sampling methods for SPCNN is presented in Table 2. The random sampling results are added to the table to highlight the importance of using balanced data. The usage of the US, EUS, and WRS methods show substantial performance improvement compared to the random sampling method. The difference is up to 40 percent.

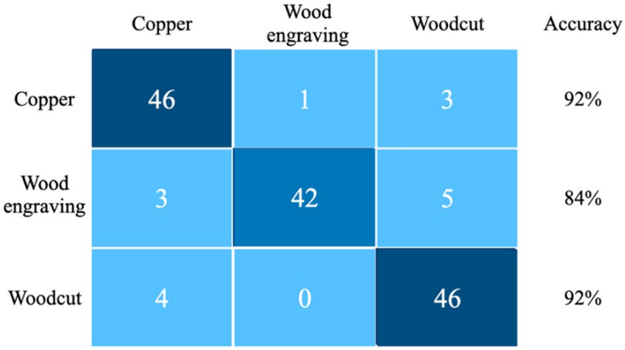
Regarding the effect of data preprocessing, all sampling methods except for EUS present better performance when D1 is used. This is related to the way EUS works. EUS guarantees that the model can learn from the entire training data. On the contrary, the other sampling methods allow that the model is not shown some training samples as they are based on probabilistic sampling. Note that the portion of outliers contained in D1 is much higher than in D2. EUC exposes our model to all images which includes outliers contained in D1.

4.3 Relation to dataset size

To inspect the interrelation of the data size and the model architectures, additional experiments were conducted using D_b which is relatively smaller than D_a. Since the main focus is to find the effect of using less well balanced data on several architectures, the sampling method EUS is applied. The results are shown in Table 3.

Table 3 Performance comparison using bigger dataset (D_b). EUS: epoch-wise undersampling, D1: original image dataset. D2: a preprocessed image dataset

Sampling	EUS	
Model/Input	D1	D2
AlexNet	77.3%	78.0%
ResNet18	77.3%	74.0%
ResNet34	73.3%	76.0%
EfficientNetB0	82.7%	82.0%
EfficientNetB3	85.3%	86.7%
EfficientNetB5	85.3%	82.7%
SPCNN	**89.3%**	87.3%



	Copper	Wood engraving	Woodcut	Accuracy
Copper	46	1	3	92%
Wood engraving	3	42	5	84%
Woodcut	4	0	46	92%

Fig. 8 Confusion matrix of the best SPCNN model that used D_b and D1 (Test Accuracy 89.3%)

The results indicate that using smaller datasets leads to better performances in all model architectures.

This finding is contrary to the general assumption about deep learning applications that they typically require large amounts of training data. The reasons for the better performances are not evident. However, it is assumed to be resulting from the stronger limitation which is imposed when creating D_b. The attempt to balance the data on the book level, in addition to balancing the classes, seems highly effective.

It can also be noted from the results that the best accuracy is achieved from a shallow network (SPCNN) architecture. The confusion matrix of the best performing model is shown in Fig. 8.

4.4 Visualizing the referred regions

Deep learning models are known for their in-transparency and lack of explainability. Nevertheless, breaking down such black box aspects has been the objective of Explainable AI [30]. The issue of explainability is relevant for many real-world tasks, although it is not sure how that can be formalized One method to understand the model suggested by literature is using the Class Activation Map (CAM) [56]. It shows the regions which the model considered as discriminative for the classification for this instance.

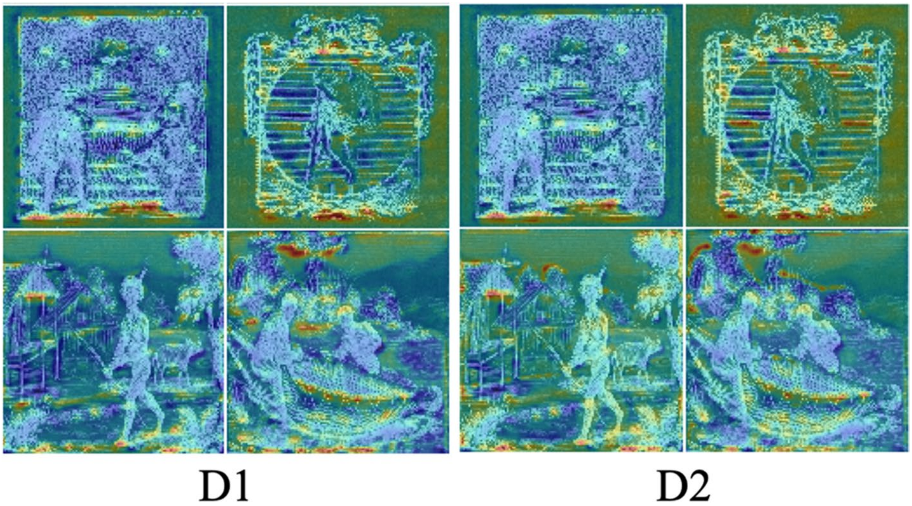


Fig. 9 Comparison of CAM images between D1 and D2

The determination of printing technology by human experts can be carried out in various ways and might consider various criteria. We claim that it is important to show the regions that are considered by the deep model not only to better understand the model decision but also to support the research related to the task. Thus, CAM is applied to the best performing SPCNN architecture. Note that the color spectrum represents the level of importance considered by the model. As in a heat map red means highest importance and blue means that these regions were not considered much by the system.

The first set of CAM images is shown in Fig. 9. The figure shows the CAM images resulting from the best performing SPCNN model which is trained on D1 and D2. Contrary to the expectations, both models generally focus on similar regions. The preprocessing seems to have an insignificant impact on the classification performance. However, the red regions shown on D2 are more concentrated on small regions and often, fine lines are highlighted. This can be clearly seen on the second image located at the top-right corners for each D1 and D2.

The second set of CAM images is shown in Fig. 10. This set is obtained from the best SPCNN model trained on D1. It can be seen from the images that the model is focusing on different regions for each printing type. The model tends to focus on the limited regions when inspecting the illustrations manufactured by the copper plates. The fine lines and edges of the objects are the most attentive areas. Similar tendencies are seen from the woodcut CAM images. The model mostly focuses on the fine details but particularly on the borders of illustrations. This is a typical mark seen in the illustrations produced by woodcut plates. However, wood engraving CAM images show that the model captures the basically the entire area. This includes the details as well as the background. Only a few images are classified considering small regions.

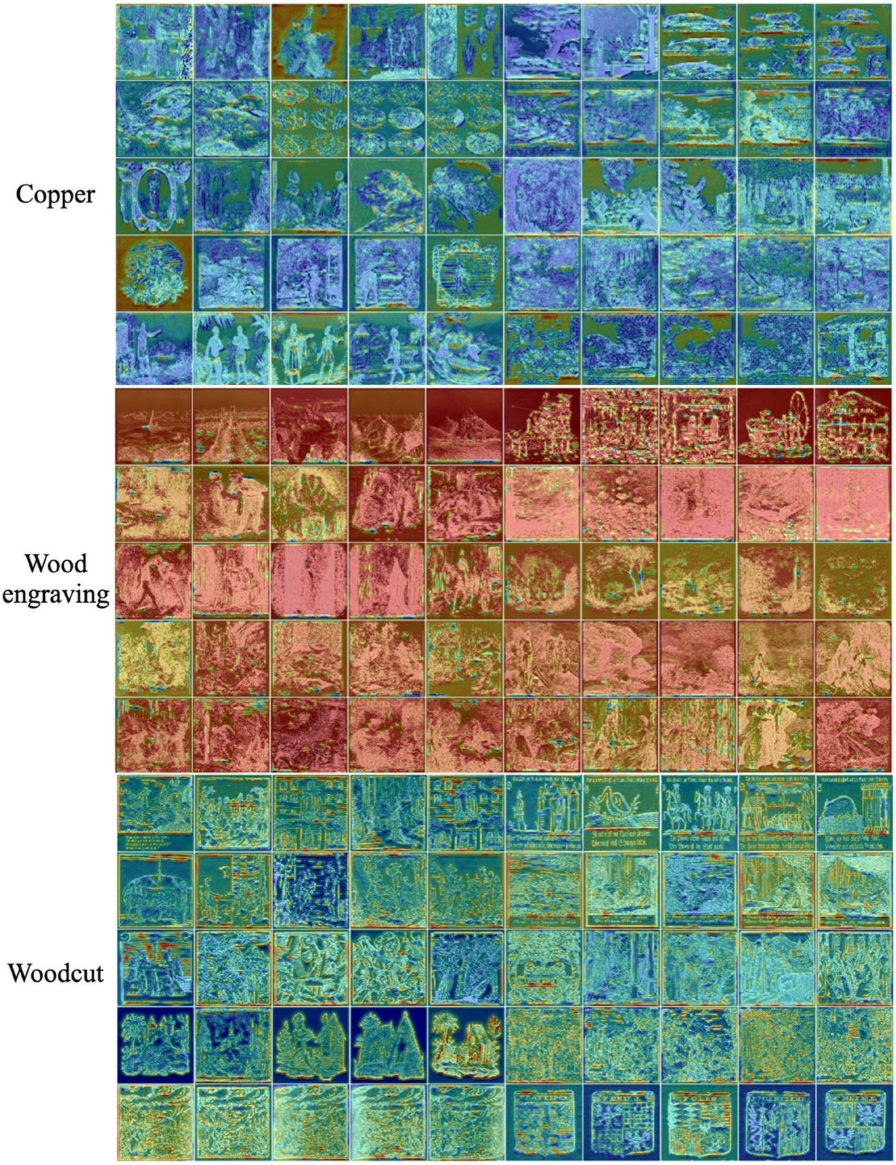


Fig. 10 CAM results of each class. The best performing SPCNN which is trained on D1 is used

5 Discussion

Our results show in general that modern computer vision architectures can be adapted to historical data as it is present in digitized book collections. Although deep learning systems are typically trained with photographs it is possible to adapt them to specific tasks like printing type identification.

The task of printing technology identification can be solved with sufficient quality. In case, the information about the printing technology is not available in digital collections it can be added automatically with a low margin of error. Deep learning architectures are even capable of this task with the limited amount of data that can be provided currently.

Shallow architecture shows better performance for this task than the deeply structured networks. This is shown by comparing the proposed SPCNN shallow model architecture to the other experimental models. This reveals that smaller patterns that are picked up in lower layers of the CNNs are of importance for detecting the printing technology.

However, for the task detailed understanding of the process is necessary. Out of the many variants, it could be shown that the sampling method needs to be carefully selected.

The operation of the networks exhibits some similarity to human behavior. It can be observed that humans often concentrate either on large areas or longer lines when trying to identify the printing type. The analysis using CAM shows that the relevant regions for making decisions lie often outside the content rich parts of the images and also in larger areas. Further studies with domain experts would be necessary to better understand their work.

6 Conclusion

In this paper, we have shown that the identification of printing technology in historical image collections can be accomplished with deep learning systems with a satisfying accuracy. We elaborated that a shallow network architecture is superior for this task because it considers smaller features.

This research opens opportunities to analyze historical image collections as included in digitized books. The printing technology can be detected automatically if the meta data is not available.

For future work, we will extend the classification task to other printing technologies and consider further collections of images.

Funding Open Access funding enabled and organized by Projekt DEAL. Fritz Thyssen Stiftung.

Availability of data and material No, copyright applies.

Code availability Yes.

References

1. Banham R (2020) The Industrialization of the Book 1800–1970. In: Simon Eliot & Jonathan Rose (ed) A Companion to the History of the Book. https://doi.org/10.1002/9781119018193.ch30.
2. Briggs A, Burke P (2009) A social history of the media: From Gutenberg to the Internet. Polity
3. Buades A, Coll B, Morel J (2005) A non-local algorithm for image denoising 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 60–65 vol. 2, https://doi.org/10.1109/CVPR.2005.38
4. Cetinic E, Lipic T, Grgic S (2019) A deep learning perspective on beauty, sentiment, and remembrance of art. IEEE Access 7:73694–73710
5. Cimpoi M, Maji S, Kokkinos I, Vedaldi A (2016) Deep filter banks for texture recognition, description, and segmentation. Int J Comput Vision 118(1):65–94
6. Crowley EJ, Zisserman A (2014) The state of the art: Object retrieval in paintings using discriminative regions. In Proceedings British Machine Vision Conference 2014. BMVA Press
7. Del Bimbo A, Pala P (1999) Shape indexing by multi-scale representation. Image Vis Comput 17(3–4):245–261
8. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: A large-scale hierarchical image database. In IEEE conference on computer vision and pattern recognition, pp 248–255
9. Donig S, Christoforaki M, Handschuh S (2016) Neoclassica-a multilingual domain ontology. In International Workshop on Computational History and Data-Driven Humanities. Springer, Cham, pp 41–53
10. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, … Houlsby N (2020) An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929
11. Gatys LA, Ecker AS, Bethge M (2016) Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2414–2423
12. Gorokhovatskyi O, Peredrii O (2018) Shallow convolutional neural networks for pattern recognition problems. In 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). IEEE, pp 459–463
13. Greeshma KV, Sreekumar K (2019) Hyperparameter optimization and regularization on Fashion-MNIST classification. Int J Recent Technol Eng 8(2):3713–3719
14. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
15. Helm W, Mandl T, Putjenter S, Schmideler S, Zellhöfer D (2019) Distant Viewing Forschung mit digitalisierten Kinderbüchern: Voraussetzungen, Ansätze und Herausforderungen. In: B.I.T.online – Zeitschrift für Bibliothek, Information und Techno¬logie. Heft 2, S. 127–134. https://www.b-i-t-online.de/heft/2019-02-index.php
16. Hossain MM, Talbert D, Ghafoor S, Kannan RR (2018) A flexible-greedy approach to find well-tuned CNN architecture for image recognition problem. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States)
17. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
18. Im C, Ghauri J, Rothman J, Mandl T (2018) Deep Learning Approaches to Classification of Production Technology for 19th Century Books. In: Lernen. Wissen. Daten. Analysen. (LWDA 2018) Workshop on "Information Retrieval" (FGIR 2018) August 22–24, Mannheim, pp 150–158. http://ceur-ws.org/Vol-2191/
19. Impett LL, Süsstrunk S (2017) From Mnemosyne to Terpsichore-the Bilderatlas after the Image. In Premiere Annual Conference of the International Alliance of Digital Humanities Organizations (DH 2017)
20. Joshi S, Saxena S, Khanna N (2020) Source printer identification from document images acquired using smartphone. arXiv preprint arXiv:2003.12602
21. Kim Y, Mandl T, Im C, Schmideler S, Helm W (2020) Applying Computer Vision Systems to Historical Book Illustrations: Challenges and First Results: In: Post-Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN) Riga. ceur_ws. http://ceur-ws.org/Vol-2865/poster7.pdf
22. Kingma DP, Ba JL (2015) ADAM: A method for stochastic optimization, International Conference on Learning Representations(ICLR) 2015
23. Kollmann S (2003) PICTURA PAEDAGOGICA ONLINE. Archives & Museum Informatics, 2
24. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105
25. Lampert CH, Mei L, Breuel TM (2006) Printing technique classification for document counterfeit detection. In International Conference on Computational Intelligence and Security (Vol. 1) IEEE, pp 639–644

26. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
27. Lehenmeier C, Burghardt M, Mischka B (2020) 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. Springer, Cham, pp 229–242
28. Lei F, Liu X, Dai Q, Ling BWK (2020) Shallow convolutional neural network for image classification. SN Appl Sci 2(1):1–8
29. Li J, Xie S, Chen Z, Liu H, Kang J, Fan Z, Li W (2020) A Shallow Convolutional Neural Network for Apple Classification. IEEE Access 8:111683–111692
30. Lipton ZC (2018) The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue 16(3):31–57
31. Luo W, Li Y, Urtasun R, Zemel R (2016) Understanding the effective receptive field in deep convolutional neural networks. In Proceedings 30th International Conference on Neural Information Processing Systems, pp 4905–4913
32. Mehri M, Héroux P, Gomez-Krämer P, Mullot R (2017) Texture feature benchmarking and evaluation for historical document image analysis. Int J Doc Anal Recognit 20(1):1–35
33. Mitera H, Im C, Mandl T, Womser-Hacker C (2021) Objekterkennung in historischen Bilderbüchern: Eine Evaluierung des Potenzials von Computer Vision Algorithmen. In: BildWissen – KinderBuch: Historische Sachliteratur für Kinder und Jugendliche und ihre digitale Analyse https://doi.org/10.1007/978-3-476-05758-7_9
34. Moretti F (2013) Distant Reading. Verso Books
35. Mustalish RA (1997) The development of photomechanical printing processes in the late 19th century. In Topics in photographic preservation: volume seven, pp 73–87
36. Neudecker C, Baierer K, Federbusch M, Boenig M, Würzner KM, Hartmann V, Herrmann E (2019) OCR-D: An end-to-end open source OCR framework for historical printed documents. In Proceedings 3rd International Conference on Digital Access to Textual Cultural Heritage, pp 53–58
37. Ramachandran P, Parmar N, Vaswani A, Bello I, Levskaya A, Shlens J (2019) Stand-alone self-attention in vision models In Proceedings 32nd International Conference on Neural Information Processing Systems
38. Saleh B, Elgammal A (2015) Large-scale classification of fine-art paintings: Learning the right metric on the right feature. Int J Digit Art Hist (2)
39. Sandoval C, Pirogova E, Lech M (2019) Two-stage deep learning approach to the classification of fine-art paintings. IEEE Access 7:41770–41781
40. Schreyer M, Schulze C, Stahl A, Effelsberg W (2009) Intelligent Printing Technique Recognition and Photocopy Detection for Forensic Document Examination. In Informatiktage (Vol. 8), pp 39–42
41. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
42. Singh A, Purohit A (2015) A Survey on Methods for Solving Data Imbalance Problem for Classification. International Journal of Computer Applications (0975–8887) Volume 127 - No.15
43. Skansi S (2018) Introduction to Deep Learning: from logical calculus to artificial intelligence. Springer
44. Strezoski G, Worring M (2018) Omniart: a large-scale artistic benchmark. ACM Trans Multimed Comput Commun Appl 14(4):1–21
45. Su D, Zhang H, Chen H, Yi J, Chen PY, Gao Y (2018) Is Robustness the Cost of Accuracy?--A Comprehensive Study on the Robustness of 18 Deep Image Classification Models. In Proceedings of the European Conference on Computer Vision (ECCV), pp 631–648
46. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, … Rabinovich A (2015) Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
47. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence
48. Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning. PMLR, pp 6105–6114
49. Tay Y, Dehghani M, Gupta J, Bahri D, Aribandi V, Qin Z, Metzler D (2021) Are Pre-trained Convolutions Better than Pre-trained Transformers?. arXiv preprint arXiv:2105.03322
50. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H (2020) Training data-efficient image transformers & distillation through attention. arXiv preprint arXiv:2012.12877
51. Van Vliet R (2019) Print and Public in Europe 1600–1800. In: Simon Eliot & Jonathan Rose (ed) A Companion to the History of the Book. https://doi.org/10.1002/9781119018193.ch28
52. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, … Polosukhin I (2017) Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pp 6000–6010
53. Wang M, Deng W (2018) Deep visual domain adaptation: A survey. Neurocomputing 312:135–153

54. Wevers M, Smits T (2020) The visual digital turn: Using neural networks to study historical images. Digit Scholarsh Humanit 35(1):194–207
55. Yang S, Oh BM, Merchant D, Howe B, West J (2018) Classifying digitized art type and time period. In Proceedings 1st Workshop on Data Science for Digital Art History-Tacking Big Data
56. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2921–2929