

Beitrag aus:

Zeitschrift für digitale Geisteswissenschaften

Titel:

Fine-Tuning Machine Learning with Historical Data. An Alchemical Object Detection Dataset for Early Modern Scientific Illustrations

Autor*in:

Sarah Lang

Kontakt: contact.sarah.a.lang@gmail.com

Institution: Universität Graz

GND: [1177417928](#) ORCID: [0000-0002-4618-9481](#)

DOI des Beitrags:

[10.17175/2025_002](#)


Nachweis im OPAC der Herzog August Bibliothek:

[1918720215](#)

Erstveröffentlichung:

27.02.2025

Lizenz:

Sofern nicht anders angegeben 

Letzte Überprüfung aller Verweise:

19.02.2025

Format:

PDF ohne Paginierung, Lesefassung

GND-Verschlagwortung:

[Alchemie](#) | [Druckgrafik](#) | [Laborgerät](#) | [Maschinelles Sehen](#) | [Technikgeschichte](#)

Empfohlene Zitierweise:

Sarah Lang: Fine-Tuning Machine Learning with Historical Data. An Alchemical Object Detection Dataset for Early Modern Scientific Illustrations. In: Zeitschrift für digitale Geisteswissenschaften 10 (2025). 27.02.2025. HTML / XML / PDF. DOI: [10.17175/2025_002](#)

Sarah Lang

Fine-Tuning Machine Learning with Historical Data. An Alchemical Object Detection Dataset for Early Modern Scientific Illustrations

- Dataset: *An Alchemical Object Detection Dataset for Early Modern Scientific Illustrations*
- Contributor: Sarah Lang (University of Graz – [Conceptualization](#) | [Data curation](#) | [Methodology](#) | [Writing – original draft](#) | [Writing – review & editing](#))
- Version: 1
- First published: 11.12.2024
- Last updated: 11.12.2024
- License: [Attribution 4.0 International \(CC BY 4.0\)](#)
- Repository: [HAB-Repositorium](#)
- DOI: [10.60831/mhk6-0r88](#)
- Suggested citation: Sarah Lang: An Alchemical Object Detection Dataset for Early Modern Scientific Illustrations. 11.12.2024. HAB-Repositorium. DOI: 10.60831/mhk6-0r88

Related publications:

- Sarah Lang: How to Create Your Own Fine-Tuning or Training Dataset for Computer Vision Using Supervisely. In: The LaTeX Ninja Blog. 06.09.2024. [[online](#)]
- Sarah A. Lang / Bernhard Liebl / Manuel Burghardt: Toward a Computational Historiography of Alchemy: Challenges and Obstacles of Object Detection for Historical Illustrations of Mining, Metallurgy, and Distillation in 16th–17th Century Print. In: Artjoms Šeļa / Fotis Jannidis / Iza Romanowska (eds.): Proceedings of the Computational Humanities Research Conference 2023 (CHR 2023, Paris, 06. –08.12.2023). Aachen 2023, pp. 29–48. PDF. [[online](#)]

1. Introduction

Annotation is a core task in the Humanities. Annotation tasks remain central to Digital Humanities research and may even gain importance as computational approaches become more widespread in the Digital Humanities. As many authors in the field have demonstrated, data work is a crucial hermeneutic task by which historical sources are modeled and transformed into data.¹ In the context of the Computational Humanities such data work is critical, as the data significantly influences the outcomes of machine learning algorithms applied to research questions. The Digital and Computational Humanities require Humanities-specific high-quality ground truth datasets curated and annotated by experts. Adapting algorithms from computer science to historical data should be straightforward in theory, but their lack of specialization for historical material – such as early modern etchings instead of contemporary photographs – poses significant challenges. Annotation specifically for machine learning applications, an essential yet labor-intensive step, is a characteristically under-resourced task.

[1]

¹ Cf. Alvarado 2022.

This article examines the challenges and considerations involved in making pixel-level annotations to fine-tune a computer vision (object detection) algorithm for recognising alchemical laboratory apparatus in early modern printed manuals. While it might appear that fine-tuning one of the currently available high-performance object detection algorithms to work with Humanities data should be an effortless task, earlier studies indicate that these algorithms encounter more difficulties with early modern etchings than initially expected: at first, a smaller set of images was annotated, featuring pictures and illustrations of early modern alchemical and chymical laboratory apparatus related to metallurgy, mining, and distillation.² The aim was for the algorithm to detect, segment, and classify these objects. [2]

This study focuses on the annotation of an additional corpus of images to increase the dataset for re-training the model. This approach was adopted after earlier attempts revealed that zero-, one-, or few-shot learning methods – which initially seemed ideal for handling data that modern algorithms have not been trained on – failed to produce satisfactory results. This outcome is not entirely surprising, as algorithms are unfamiliar with historical alchemical objects, both in terms of visual features and nomenclature. The consideration was that annotating a training set to meet the ideal sample size per category, such as 1,500 examples for a *YOLO* class (*You Only Look Once*³), was not only deemed unnecessarily labor-intensive but also, in some cases, impractical. This is because such a large number of instances for an object class may not exist in historical sources. Moreover, if all examples were already included in the training dataset, there would be no additional data to apply the algorithm to, rendering the automation effort pointless. Thus, the first attempts relied on using only a few annotated examples. Moving forward, the next step will be to determine whether increasing the size of the training set for at least one object category improves performance significantly. Alternatively, future efforts may focus on identifying whether segmentation (i. e., locating objects on the page) or classification (i. e., correctly labeling or naming the objects) is the primary source of error. This could lead to the use of algorithms that treat segmentation and classification as separate tasks, with specialized models employed for each. [3]

However, to move in this direction, more and better training data is required. It is for this reason that this article focuses on the issue of annotating additional images to extend the ground truth dataset for fine-tuning the computer vision algorithm, providing both methodological considerations and lessons learned related to this data work.⁴ Ultimately, it presents the resulting dataset, encompassing all relevant contexts, including the motivation for the work, the current literature on distant viewing in Computational Humanities, the historical genre, and the creation and technical description of the dataset. It concludes with a discussion of annotation, challenges, and the impact of Humanities-type data work on the outcomes of machine learning applications in Computational Humanities. [4]

1.1 Motivation

This dataset originates from a project aimed at applying object detection algorithms to alchemical illustrations.⁵ Initial efforts focused on few-shot approaches – particularly one-shot and zero-shot methods – assuming that these would be well-suited for historical sources. The reasoning was that adapting powerful existing algorithms for unfamiliar datasets would enable effective detection of alchemical objects. However, this approach proved less effective than anticipated: attempts to train *YOLO* with minimal annotations revealed several challenges. The algorithm struggled, likely due to the limited training data, and discrepancies between the training and evaluation splits may have further hindered performance. While some classes, such as visually distinct jars (*ollae* class), were learned successfully even with minimal examples, other classes with [5]

² Cf. Lang et al. 2023.

³ Redmon et al. 2015.

⁴ Cf. GitHub for additional documentation: Lang 2024a; blog post on the annotation process in the Supervisely platform: Lang 2024b.

⁵ Cf. Lang et al. 2023.

greater variability were more problematic. This underscores the need for more robust datasets tailored to historical sources. To address this need, this follow-up project drew on data from Herzog August Bibliothek Wolfenbüttel (HAB) to expand the training set.

Fortunately, the project had anticipated the potential future need to expand the training dataset and designed a classification scheme aligned with the alchemy taxonomy developed by Ute Frietsch using [Iconclass](#), an international classification method for visual content. In addition to classifying and tagging textual sources, the alchemy portal of the Herzog August Bibliothek Wolfenbüttel created in the project [Erschließung alchemischer Quellen in der Herzog August Bibliothek](#) (2014–2017) aimed to provide a system for tagging visual material from alchemical prints in collaboration with the Foto Marburg image archive using Iconclass. The goal was to provide users with an image index accessible via the project's online portal, enabling image searches by motif rather than by the book containing the image. The image corpus, catalogued based on Iconclass, was selected according to representative aspects from digitized materials. The classification aimed to cover a broad spectrum of image types, ranging from technological depictions to allegories and illustrations of pharmaceutical plants and animals, as these often accompany pharmacological aspects in alchemical works. Through the collaboration with Foto Marburg, it was possible to extend the Iconclass framework in aspects where it had proved inadequate for certain alchemical material. Notably, numerous new terms related to technical apparatus and alchemical processes were introduced to the class [49E39 ›alchemy‹](#). The approximately 3,000 images digitized and catalogued during the project can be cross-referenced with images from other collections through their aggregation in databases like [Virtuelles Kupferstichkabinett](#) and the Dutch image archive [Arkyves.org](#), allowing for searches by Iconclass code, subject or keyword.

[6]

1.2 On the historical source material

Focusing on visibility has not only been a trend in the recent Computational Humanities paradigm of ›Distant Viewing‹ but has become a prominent topic in the historiography of alchemy as well.⁶ Due to the scarcity of material evidence from early modern alchemical laboratories – only a few having been discovered and excavated – the combination of visuals and text in early modern handbooks and technical manuals serves as a vital source for understanding the laboratory operations of early modern alchemy and chymistry.⁷ Despite their significance in the history of technology, these books had long been relatively understudied. However, in recent years, there has been a growing trend to investigate these sources, particularly regarding their visual elements and depictions of laboratory apparatus. They offer unparalleled insights into the laboratory practices and processes underlying the *artes technicae* of the period.

[7]

During the Proto-Industrial Revolution⁸, a period of accelerated industrial development in the 16th–19th centuries leading up to the Industrial Revolution, flourishing mining and metallurgy gave rise to encyclopedic compendia detailing technological apparatus and processes. Metallurgical and distillation treatises became staples of didactic manuals and were often accompanied by technical illustrations. Beginning with smaller treatises, works of this genre evolved into larger and more comprehensive compendia by the mid-16th century. These books were particularly valuable as they partly replaced the need for direct exchanges with experts and eliminated the necessity of costly and time-consuming journeys, even though not all techniques they described were equally practical or easily implemented. Such metallurgical and distillation handbooks were widely published, translated, and reissued in various editions. Despite their importance for the history of technology and everyday knowledge, handbooks, manuals, and pragmatic literature still remain relatively underexplored in scholarly literature.

[8]

⁶ Cf. Götzelmann 2022; Laube 2022; Wels 2022; Hofmeier 2023; Purš / Karpenko 2023. On distant viewing, see Arnold / Tilton 2019.

⁷ Cf. Lang 2023.

⁸ Cf. Mendels 1972.

1.3 Distant Viewing Illustrations in Early Modern Alchemical Print

This effort contributes to the broader trend of applying computer vision to Digital Humanities, referred to as *distant viewing* in analogy to the established concept of *distant reading*.⁹ While distant viewing is often applied to modern datasets, such as video data that are much closer to the type of data that modern algorithms are trained on, its application to historical datasets is more complex and necessitates reliable training datasets to improve algorithm performance. [9]

The increased availability and reuse of illustrations contributed to the emergence of large, elaborately illustrated encyclopedias in the late 17th and early 18th centuries. A well-trained algorithm capable of recognizing specific objects or similar illustrations would enable researchers to trace the evolution of visual practices and their role in the history of science and practical knowledge. Such an algorithm could reveal trends in realism, technological advancements, and the transmission of knowledge through images. The algorithms used so far, such as the Oxford Visual Geometry Group's *VISE*, rely on methods like *keypoint extraction* to identify similar objects. This approach has been used effectively by a group around the book historian Giles Bergel to track reused illustrations in early modern datasets.¹⁰ [10]

For instance, early botanical treatises, where visual accuracy was crucial for sales, frequently experienced copyright disputes over stolen illustrations.¹¹ As illustration plates were expensive, printers frequently reused them, creating a system akin to modern stock photography. This reuse likely weakened the match between text and image, although this remains a hypothesis due to the absence of large-scale analyses. Scholars like Germaine Götzelmann have explored illustration similarity detection – a related but more established method in Digital Humanities approaches to early modern print culture. Recent work by the [MPIWG research group on De Sphaera](#), an early astronomy textbook, used explainable AI to study the reuse and dissemination of its illustrations across editions, providing insights into the spread of astronomical knowledge.¹² Object detection, though distinct from similarity detection, shares methodological connections and holds immense potential for uncovering visual trends in early modern science and publishing. This dataset represents a valuable resource for advancing these efforts, enabling scholars to deepen their understanding of early modern visuality, scientific practices, and the interplay between text and image. [11]

2. A Computer Vision Dataset for Alchemical Illustrations: Data Collection

This annotation data has its roots in an image dataset curated by Ute Frietsch at the HAB as part of a previous [project](#). In this project, Ute Frietsch created Iconclass classifications for alchemical objects.¹³ This was accompanied by images from the Wolfenbüttel Digital Library, which included alchemical laboratory objects. As part of this effort, the books were digitized and made available online. Detailed descriptions are accessible in the OPAC, although, unfortunately, they cannot be easily queried or downloaded outside the OPAC. In the project *Erschließung alchemiegeschichtlicher Quellen* at the HAB, a dataset of 1,800 relevant book pages from early modern prints was tagged with keywords. This was done using a classification scheme based on Iconclass, aligned with the content of the included illustrations.¹⁴ However, this classification applies only to entire book pages, which often contain more than just the relevant illustrations, such as columns of text. Furthermore, one illustration usually shows more than one alchemical object. So, while this dataset holds promise for machine learning applications such as computer vision, it was not yet directly usable in its [12]

⁹ Cf. Arnold / Tilton 2019.

¹⁰ Cf. Dutta et al. 2021.

¹¹ Cf. Götzelmann 2022.

¹² Cf. Büttner et al. 2022; El-Hajj et al. 2023; Valleriani et al. 2023.

¹³ Frietsch 2017.

¹⁴ [Image browser at Arkyves.org](#).

current form. Most algorithms, particularly those for object detection, require precise information about the location of objects on a page. To make the data usable for computer vision and specifically object detection, the locations of the objects on the images need to be specified in a suitable image annotation format (such as *MS COCO*¹⁵) so they can be used to train machine learning algorithms. Furthermore, many algorithms do not support multiple annotations per page, especially if they overlap within the image. However, many images from the Herzog August Bibliothek include multiple tags.

In the context of a [Wolfenbüttel NFDI 4 Memory Fair Data Fellowship](#), approximately 640 of these pages were classified using [the Supervisely platform](#).¹⁶ The images used to create this dataset are individual pages from the HAB collections that had already been tagged with keywords. Based on these tags, a script was used to retrieve relevant images. The Supervisely platform's annotation tool was used to create pixel-level annotations for the alchemical objects in these images adapting the classification proposed by Ute Frietsch. However, the images themselves are not the dataset; it is rather the annotations that constitute the dataset discussed here, as the images had already been published in their own place. [13]

To make the resulting dataset easier to analyze, [two scripts](#) for scraping the relevant images and processing metadata for the books included in the dataset from HAB's digital library were created. It is important to have an overview of which books and images are included in the dataset, for example so as not to mix up the training and evaluation data. The metadata extraction script will generate a new CSV file containing the following fields: the original image link from the input, the viewer link, the link to the METS XML file, the full bibliographic citation string, and the separate fields such as author, title, publishing place, publisher and publication year, as far as it was possible to automatically extract them from the citation string as the METS file this is retrieved from only contains an unstructured bibliographical reference. [14]

3. Technical Dataset Description

This dataset contains pixel-level annotations for 12 categories of alchemical objects derived from images of digitized manuscripts held at the HAB. The annotations are provided in two formats, *YOLOv8* and *MS COCO*. The dataset is designed for object detection tasks relating to apparatus depicted in early modern manuals of chymistry, metallurgy, mining and distillation. It includes categories such as *ampullae*, *cucurbitae*, furnaces, and more.¹⁷ An example of what annotating the images looked like in the Supervisely platform can be found in the accompanying blog post.¹⁸ [15]

¹⁵ Lin et al. 2014.

¹⁶ For details, see Lang 2024b.

¹⁷ The annotations cover 12 categories of alchemical objects: *ampullae*, animal, *cucurbitae*, *cucurbitae-ambix*, *cucurbitae-retorte*, *cucurbitae-roshut*, furnace, human, mineral-metal, *ollae*, other-equipment, plant.

¹⁸ Lang 2024b. The time required for annotating images varies significantly depending on the complexity of the image. A simple image with a single object and no background can be annotated in as little as 10 seconds. However, annotation time increases proportionally with the complexity of composite setups, which can be considerable. In many cases, [Supervisely's »Smart Tool«](#) (»magic wand«) accurately identifies object boundaries, but when the tool struggles to distinguish between objects in certain images, time demands escalate sharply. This not only extends the time required for annotation but also depletes the daily limit of »smart points« available in the free version, as their number is capped per day. Further delays occur when it is unclear how to label an object, requiring the annotator to refer back to the accompanying text for clarification. Even then, this process is not always straightforward, as not all illustrations are accompanied by detailed explanations in the source material. However, using *Segment Anything (SAM) 2* as underlying algorithm was a significant improvement compared to the older models. In general, the tool performs optimally with high-resolution images featuring relatively simple structures, as is often the case with early modern illustrations resembling blueprints, technical drawings or simply sketches. Conversely, more complex and less sketch-like images require significantly more effort to annotate using the Supervisely tool. In these instances, automatic segmentation may struggle, and a large number of smart points may be required to refine the annotations. In these cases, the otherwise highly »intelligent« SAM2 models can be stubborn and resistant to adjustments. In such instances, it may be more effective or even necessary to switch back to older models. However, SAM2 performs significantly better for images featuring overlapping or pixelated objects. It excels in handling images with depth of space, regardless of how this is achieved artistically, whereas older models gave up when faced with excessive cross-hatching. This capability allows SAM2 to annotate objects in the background, a task that was almost impossible to accomplish semi-automatically with earlier models. In practice, using Supervisely's magic wand tool involves first drawing a rough bounding box around the object. The tool then suggests a segmentation, which the user can refine by placing

The following snippet from the MS COCO file demonstrates the structure of the data. In ›segmentation‹, points demarcate the object's boundaries. As seen in some examples in the dataset, these points can be numerous, depending on the object's level of detail. The bounding box (›bbox‹), on the other hand, is a rectangle outlining the object's location, defined by four corners. Each annotation receives an ID number, while the category ID specifies the class to which it belongs. At the bottom of the example are the different categories (grouped as supercategories). Each category has an ID number and a name, although this name is what we primarily interact with in practice.¹⁹

[16]

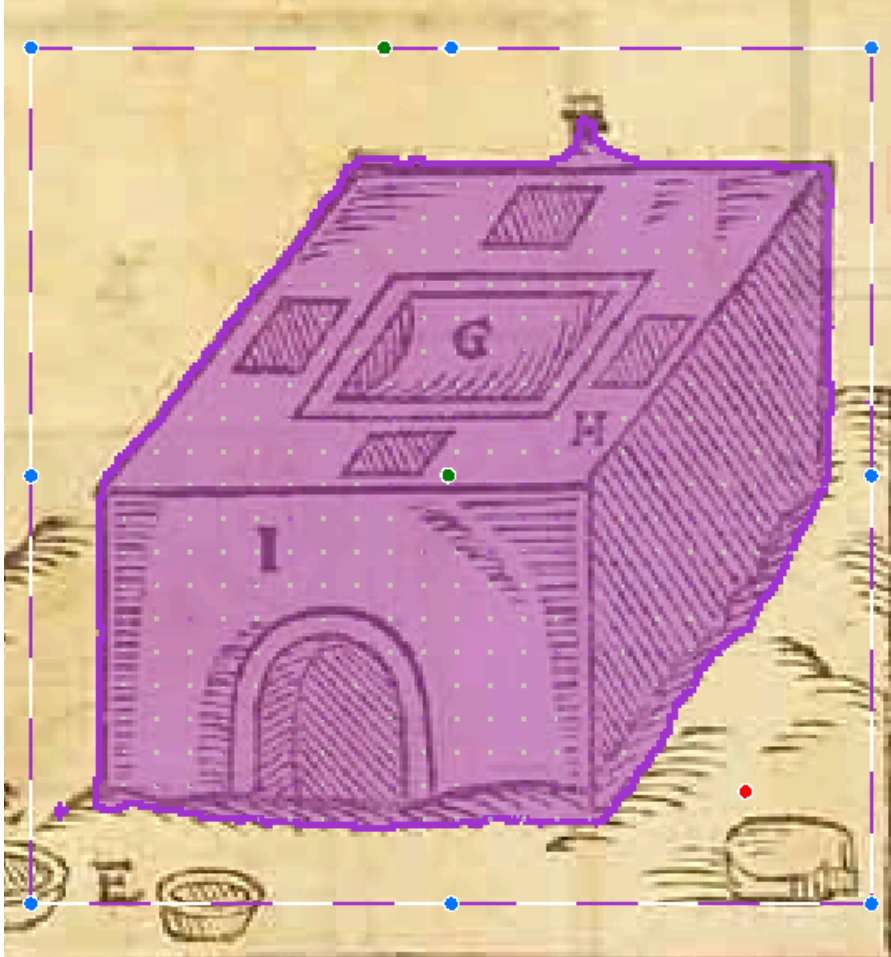


Figure 1: Illustration of a furnace from Andreas Libavius' *Alchymia* (Frankfurt 1606), overlaid with segmentation and boundary box. [Source: Supervisely / WDB (Public Domain)]

```
1. {
2.   "info": {
3.     "description": "This dataset contains pixel-level annotations (segmentation
plus label) for 12 classes of alchemical objects using data from HAB
Wolfenbüttel.",
4.     "url": "https://github.com/sarahalang/hab-nfdi4memory-fairDataFellow",
5.     "version": "1.0",
```

smart points. Positive points, indicating areas that should be included in the annotation, are added by clicking, while negative points, indicating areas to exclude, are set by holding the Shift key while clicking. The advantage of this method is that in addition to a bounding box, this process adds pixel-level information about the object's location to the annotation data.

¹⁹ In the image information, ›date_captured‹ refers to the upload date in the Supervisely app. The ›date_created‹ refers to the date the dataset was exported in MS COCO format from the Supervisely project. The example refers to a furnace object in an image from Andreas Libavius' *Alchymia* (Frankfurt 1606), cf. Fig. 1.

```
6.         "year": 2024,
7.         "contributor": "Sarah Lang",
8.         "date_created": "2024-09-25T10:01:27.822Z"
9.     },
10.    "licenses": [
11.        {
12.            "url": "None",
13.            "id": 0,
14.            "name": "None"
15.        }
16.    ],
17.    "images": [
18.        "license": "None",
19.        "file_name": "nd-4f-18_00368.jpg",
20.        "url": "None",
21.        "height": 1730,
22.        "width": 1024,
23.        "date_captured": "2024-09-25T10:03:20.911Z",
24.        "id": 342954421
25.    ]
26. ],
27. "annotations": [
28.     "segmentation": [
29.[
30.         702.0,
31.         1103.0,
32.         702.0,
33.         1107.0,
34.         704.0,
35.         1105.0
36.     ]
37. ],
38.     "area": 4.0,
39.     "iscrowd": 0,
40.     "image_id": 342954421,
41.     "bbox": [
42.         702.0,
43.         1103.0,
44.         2.0,
45.         4.0
46.     ],
```



```
47.         "category_id": 8,
48.         "id": 6979
49.     },...{
50.         "supercategory": "cucurbitae-rosenhut",
51.         "id": 7,
52.         "name": "cucurbitae-rosenhut"
53.     },
54.     {
55.         "supercategory": "furnace",
56.         "id": 8,
57.         "name": "furnace"
58.     },
59.     {
60.         "supercategory": "human",
61.         "id": 9,
62.         "name": "human"
63.     },
```

In the YOLO data format, the structure differs slightly, but image annotation formats ultimately all provide roughly the same type of information, helping an algorithm locate specific areas on a page – our annotated pixel-level data. The bounding box is particularly important; for instance, it allows parts of an image to be cropped. Some algorithms only accept one annotation per image, whereas our images typically require multiple annotations. In our earlier paper,²⁰ we used this information to extract image segments to feed directly into an algorithm, effectively transforming the object detection task into a simpler image classification task. This demonstrates one practical application of a bounding box. Segmentation, on the other hand, helps the algorithm distinguish the relevant area from the background. In this sense, the bounding box provides a general location, including background elements, whereas the pixel-level segmentation isolates only the relevant pixels. For effective algorithm training, the more precise the information, the better – making pixel-level annotation particularly valuable.

[17]

Each image's YOLOv8 annotations are stored in .txt files within the train/ directory.²¹ To use these annotations with a YOLOv8 model, ensure that your image dataset is organized accordingly. The annotations contain coordinates normalized relative to the image size that the model can use to detect objects during training or inference, making them model-ready for YOLOv8. The annotations within each .txt file follow the YOLOv8 format, where each line represents an object to be detected in the image with the following structure:[class_id] [x_center] [y_center] [width] [height] [x1] [y1] ... [xN] [yN]. In the yolo8-labels/train/ directory, each image has a corresponding .txt file with its annotations. These .txt files are named according to the pattern: [Supervisely Prefix]_[HAB Signature]_[Page Number].txt. For example: 971588_nd-2f-1-1b_00973.txt. The 971588_ is an irrelevant prefix, automatically generated by the Supervisely platform, while the remainder of the filename (nd-2f-1-1b_00973.txt) reflects the manuscript's

[18]

²⁰ Lang et al. 2023.

²¹ Cf. [YOLOv8 Label Format](#).

call number (nd-2f-1-1b) and the corresponding page number (00973). To reference an image based on its YOLO annotation file name, simply use the HAB call number and page number found in the name of the .txt file. These can be matched with entries in the accompanying .csv files to find relevant links and metadata.²²

MS COCO annotations provide information for object detection and segmentation tasks and are stored in a single JSON file, `hab-alchemieobjekte-msCoco.json`.²³ This format is more human-readable and provides detailed information about images, objects, and segmentation masks with components like general dataset information, including description, contributor, and version, metadata for each image, such as file name, dimensions, and an ID; the annotations, i. e. object annotations for each image, including segmentation polygons, bounding boxes, category IDs, and object IDs. Finally, the file lists the 12 object categories with corresponding IDs and names. To use these with a COCO-compatible model, load the `hab-alchemieobjekte-msCoco.json` file and ensure your images are structured in accordance with COCO conventions. [19]

The actual image files are not included in this repository. However, you can download the images using the provided scraping script, which fetches them from the Herzog August Bibliothek's digital library. The .csv files included in the dataset contain download links and metadata that help locate the corresponding images and books within the library's online collection. Included with the dataset are .csv files relating to the metadata for images from the first dataset used in this project that was compiled from the Austrian National Library's (ÖNB) Digital Library (`CHR2023-corpus-train.csv` and `CHR2023-corpus-eval.csv`). This information would be necessary if you wanted to recreate that image corpus to use the annotations of the full dataset in `ms-coco-full-dataset.zip` and `yolo-train-full-dataset.zip` or investigate if the resulting complete dataset requires further deduplication due to images from the same books being present multiple times. [20]

4. Annotating Alchemy and the Alchemy of Annotation: Lessons learned on the influence of annotation in creating ground truth datasets for machine learning applications in the Computational Humanities

One of the primary issues encountered involves the visual similarity between various types of alchemical vessels, such as *cucurbitae* and *ampullae*. These objects, though different in their purpose of use, are so alike visually that it is doubtful whether an algorithm can reliably differentiate between them (cf. Fig. 2).²⁴ [21]

²² `hab-alchemieobjekte-msCoco.json` contains the object annotations in MS COCO format, which are more human-readable than the YOLO format. `yolo8-labels-train.zip` is the `train/` folder of YOLOv8 formatted annotations. `yolo8-labels-data_config.yaml` is the configuration file for the YOLOv8 annotations. The configuration file lists the annotation classes, color information, and paths to the train and val datasets. Although paths to these datasets are specified, you need to download the images yourself: they are not included in this dataset as they are already archived in HAB's digital library. Look up how to properly structure your data so it's ready for actually using the algorithms. `yolo-train-full-dataset.zip` contains the full dataset (not just the data from HAB but also from the books in Lang et al. 2023). The books used are cited there including the link to the facsimiles we used and the images are named accordingly but it will likely be more work to assemble all the required images. To make it easier for you to write a download script to get those books and images, I have included the two csv files `CHR2023-corpus-train.csv` and `CHR2023-corpus-eval.csv` with the relevant metadata.

²³ You can learn more about the MS COCO dataset and format here: [COCO dataset official site](#) or [Papers with Code – COCO dataset introduction](#). Despite being filled mostly with numbers, these data formats are relatively easy to understand. Each image contains all annotations within a single level of nesting. For each annotation, there is a bounding box (a rectangular box defining the object's location), pixel-level segmentation (coordinates outlining the object), and the associated class tag. Additionally, the image filename is typically included. A Python script can easily extract relevant information. For example, the bounding box coordinates can be used to crop sub-images for classification training, as was done in the initial project. The class tags can be used to sort images into folders based on their category, a common setup for training computer vision algorithms. In this dataset, this process would still need to be done manually, as the images were not pre-organized by class and often contained multiple classes. The setup would involve extracting the relevant image snippets, determining their class, and placing them into folders named after the corresponding class. Alternatively, the data could be arranged into a CSV list or DataFrame for training, as many computer vision algorithms can process data in this format as well.

²⁴ In Fig. 2, the image on the left, for instance, highlights the challenge of annotating images based on a concept-driven rather than a purely visual classification. It depicts two visually similar vessels that serve different functions, leading to different annotations: one as a flask-like device, a *receptaculum*, and the other, placed in the furnace, as a *cucurbit*. Despite their resemblance, their functional differences dictate their names. A similar issue appears in the bottom-most illustration of the right image, where two *cucurbits* are connected to a smaller container, which functions as a *receptaculum* (a receiver vessel), thus annotated as a flask-like

Even for humans, distinguishing these vessels is difficult; their simple shapes often require contextual clues for identification. *Cucurbitae*, for instance, are marginally rounder, resembling pumpkins, but such subtle distinctions may not always be apparent in the illustrations. Moreover, the objects are frequently labelled and described in the accompanying text, given the instructional nature of the books. However, even this does not always simplify the process: terminological inconsistencies across texts present another challenge. While »cucurbita« is used in both Latin and German, authors sometimes employ more general terms (e. g., »vas« or »Gefäß«), and the precision of descriptions varies.²⁵ The issue becomes more pronounced when distinguishing between vessels used in similar contexts. Receiver vessels, classified as *ampullae* by Ute Frietsch, are difficult to distinguish visually from slender *cucurbitae*, as they often appear in similar positions within experimental setups. Consequently, it remains uncertain whether drawing a strict distinction between similar and relatively general objects like *ampullae* and *cucurbitae* is even meaningful, especially for such less differentiated objects. For more distinctive items, such as distillation helmets, retorts, or *rosenhut* devices, this is not an issue, as their unique features make them easier to categorize.

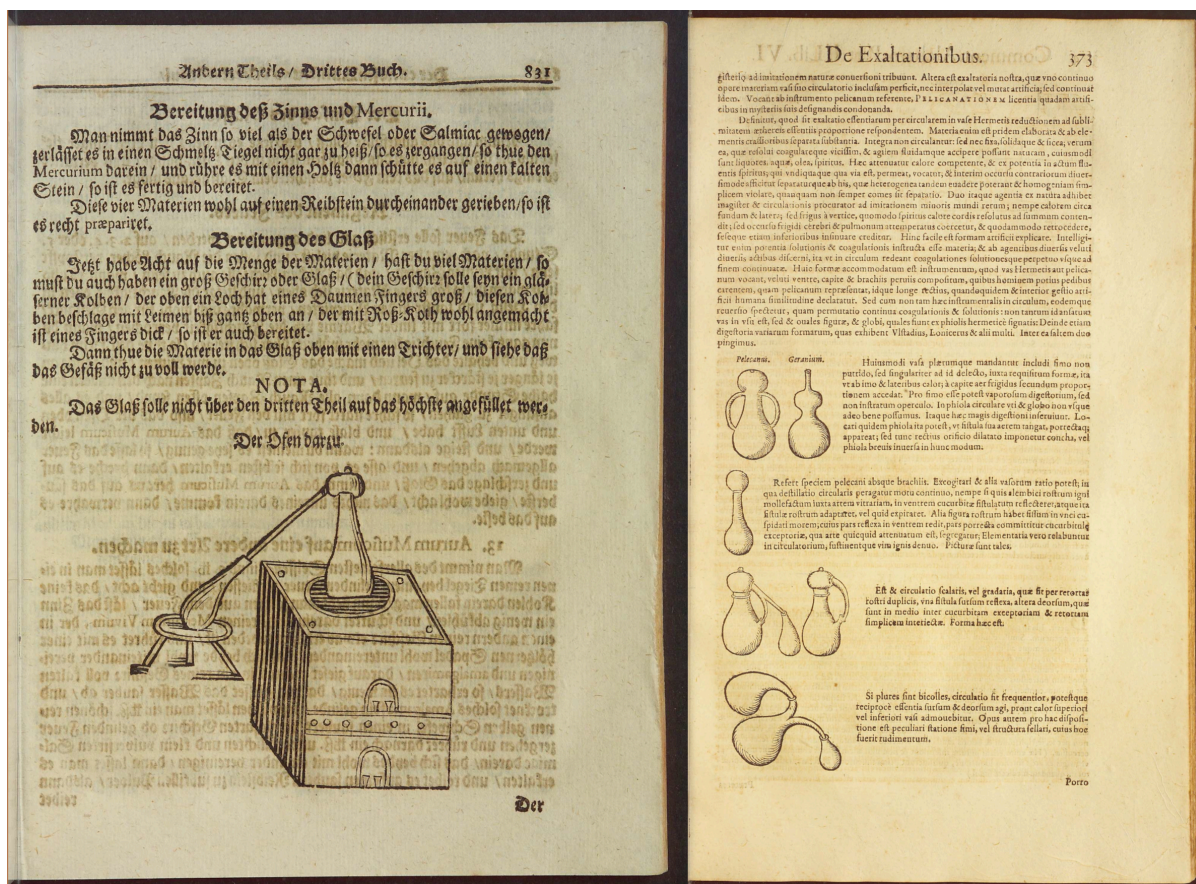


Figure 2: Illustrations of various alchemical vessels from Johann Kunckel's *Der curieuses Kunst- und Werck-Schul Anderer Theil* (Nürnberg 1707, left) and Andreas Libavius' *Alchymia* (Frankfurt 1606, right). [Source: WDB, oc-77-2 / nd-4f-18 (Public Domain)]

Another complexity arises with combination devices, which present challenges both for the algorithm and for manual annotation. Due to the project's initial objective of minimizing the number of categories, certain composite devices, such as *alembics* or *moor's heads*, were not explicitly available as annotation categories. Despite their visual distinctiveness, these devices occur infrequently in the books, so they were not prioritized. However, this decision may need to be revisited as the project progresses, especially if, at

[22]

device. Although visually rounded and similar to a *cucurbit*, its classification is ultimately determined by its function in the setup, both in Ute Frietsch's Iconclass classification and the alchemical terminology (cf. Gaede 2024) it was based on, rather than by its visuality that is more relevant for machine learning applications.

²⁵ Similar observations have recently been made in a terminological study (Gaede 2024).

a later stage, these devices are deemed important enough to warrant their own categories. At this stage, though, adding more categories would have only further complicated the process, especially given that the object detection algorithm had still struggled even with simpler categories. Fortunately, since these combination devices appear in a relatively small number of images, it should be feasible to correct the annotations later by introducing specific categories and re-annotating the relevant objects, should the need arise. In sum, we initially avoided annotating composite devices due to the limitations of the algorithm we were using. However, we later recognized that the composite and overlapping nature of these objects is a defining feature of alchemical images. Training the algorithm to handle these complexities is essential and ignoring them would ultimately undermine the entire effort. This experience has provided valuable insights into how to approach the annotation of complex objects in historical images.

A further issue concerns the inclusion of duplicate images within the training dataset, particularly due to the integration of the HAB dataset. While the first annotated data from the Austrian National Library that was the training material for the original paper²⁶ did not seem to overlap with HAB Wolfenbüttel scans, the issue arises from the possibility of different copies of the same edition being part of the HAB collections. Additionally, illustrations may have been reused across multiple early modern books, which could introduce further duplicates. This phenomenon of image reuse, common in historical works, still needs to be investigated further. Despite efforts to deduplicate the data, it appears that some images within the HAB Wolfenbüttel corpus were included more than once.²⁷ This could result in the same image being present in both the training and evaluation splits of the model training later, which would compromise the validity of the results. Now we are faced with the question whether such duplicates can be used effectively in training: if nearly identical images are placed in both the training and evaluation datasets, it may artificially inflate the algorithm's performance, as it would appear to have learned more than it actually did. On the other hand, there is an argument for retaining these duplicates, since a realistic use case may involve finding similar objects across different editions of the same book. A potential solution could involve clustering highly similar images and pruning those that are too similar, preventing the algorithm from memorizing them and hindering its ability to generalize to unseen cases. Future work should accordingly analyze and visualize the dataset and the annotations for possible duplicates and patterns in the bibliographic metadata. [23]

In the annotation process, we encountered challenges in matching the existing classification scheme with the actual data. This highlights the vital hermeneutic role of annotation and the potential complexities it introduces. Once this data is applied in machine learning, the limitations of the annotations will directly impact the algorithm's performance. Thus, the epistemic and hermeneutic considerations behind the annotation are crucial, as they influence how the algorithm learns and processes the information. [24]

One lesson learned from this project is that with the experience of annotating large numbers of images, we might have devised a slightly different classification scheme. The theoretical model developed by Ute Frietsch, while logical, was not based on direct image annotation, which created some discrepancies between the theoretical classification based on concepts and the actual visual characteristics of the images. For example, Frietsch classified receiver vessels (*receptaculum*) among flask-like objects due to their size and function, while cucurbits, typically larger and rounder, were placed in a separate category. However, these objects can look visually similar, which complicates both annotation and algorithmic classification. The issue is exacerbated by the fact that computer vision algorithms rely more on visual cues than such functional distinctions that can, if at all, only be gleaned by studying the accompanying text. This discrepancy between theoretical models and the visual characteristics of the objects also posed challenges for the human annotator. This experience with annotation underscores the importance of aligning the theoretical classification model with the actual appearance of objects in historical images. The hierarchical structure of Frietsch's classification, [25]

²⁶ Lang et al. 2023.

²⁷ Since the images are named using the format `(signature)_(page).jpg`, duplicates of the exact same image technically should not exist. However, it is possible that multiple copies of the same book were digitised and included in the HAB alchemy portal, resulting in virtually identical images. This issue requires further investigation to ensure that duplicates do not affect the training process.

which separated *receptaculum* and *cucurbita* into distinct classes despite their visual similarities, may confuse the algorithm. A more flexible classification system, perhaps with subclasses for distinct objects and a general ›flask-like device‹ category for ambiguous cases, might help the algorithm better understand the variations in these early modern illustrations.

There also was a clear tension between keeping the classification system general, as initially planned, and introducing more specific classes based on the subject matter. The original idea had been to limit the classification to five broad categories: furnaces, humans, plants, pots / jugs / bottles, and minerals / metals. However, as the project progressed, it became evident that this approach would be too simplistic, especially for distinguishing alchemical objects from non-alchemical ones. Some alchemical devices are so distinctive that they are crucial for identifying an alchemical context, which led to expanding the classification to 12 classes. [26]

Some objects, such as different types of furnaces, were initially excluded due to an assumed lack of examples, but further investigation revealed that there would have been enough instances in the (now extended) dataset to justify including them. These furnaces, which unlike the flask-like vessels can be quite visually distinct, might have been easier for the algorithm to classify than the more ambiguous flask-like objects. A future approach could involve clustering objects within the furnace class and assigning sub-labels based on their visual similarities to refine the annotation scheme to include the sub-categories after all. Additionally, certain highly specific objects, like the pelican flask, were excluded due to the limited number of examples, despite their distinctiveness. In hindsight, it may have been beneficial to include such objects, given how easy they should be for an algorithm to recognize, despite their rarity. [27]

This shift toward more specific classifications addresses the need for greater detail but introduces a new challenge: ultra-specific classes are easier to distinguish visually and build less ambiguity into the annotation process, which likely results in uneven annotations over larger datasets, but they come with the drawback of fewer examples being available, making it harder to train the algorithm effectively. We have not yet reached a definitive conclusion on how to balance the need for specificity with the practical limitations of having enough historical examples to annotate and classify, which will likely be a consideration in future work. [28]

In addition to presenting the dataset, this paper has offered a theoretical and methodological account of the impact of this type of annotation work in Computational Humanities. Such tasks often remain side notes in Computational Humanities publications, despite being crucial for the accuracy and success of the algorithms they support. The article has illustrated the complexity and the numerous sub-tasks and decisions involved in such work, even for a relatively well-defined problem. Yet this data work is often seen as menial or just a necessary ›pre-processing‹ task that comes before the actual work of applying machine learning to Humanities data, but in fact this data work is crucial for the outcomes of all machine learning. [29]

This discussion addresses a core challenge in Digital Humanities: mapping historical objects of study and their sources onto simplified structures that meet the needs of computer systems, which rely on straightforward classifications. The paper has discussed the hermeneutics of this process, commonly referred to as annotation, modelling or data work²⁸, and examined the problems and challenges that arise when balancing the needs of scholars and their research questions, the constraints and demands of historical material, and the requirements of the computer vision algorithms being used. The technical considerations and annotation challenges discussed here reflect the broader implications of integrating computational methods into Digital Humanities research, demonstrating that both algorithmic innovation and Humanities-informed data work are essential to achieving meaningful results in Computational Humanities projects. In [30]

²⁸ Cf. Alvarado 2022.

conclusion, this dataset represents a step forward in leveraging AI for the study of illustrations of alchemical laboratory objects but also demonstrates the need for iterative refinement of annotation workflows, larger datasets, and tailored approaches to align machine learning tools with the realities of historical sources.

Bibliography

- Rafael Alvarado: Datawork and the Future of Digital Humanities. In: James O'Sullivan (ed.): The Bloomsbury Handbook to the Digital Humanities (= Bloomsbury Handbooks), London 2022, pp. 361–372. [\[Nachweis im GVK\]](#)
- Taylor Arnold / Lauren Tilton: Distant Viewing: Analyzing Large Visual Corpora. In: Digital Scholarship in the Humanities 34 (2019), Supplement 1, pp. 3–16. 16.03.2019. DOI: [10.1093/llc/fqz013](#)
- Jochen Büttner / Julius Martinetz / Hassan El-Hajj / Matteo Valleriani: CorDeep and the Sacrobosco Dataset: Detection of Visual Elements in Historical Documents. In: Journal of Imaging, 8 (2022), no. 10. 15.10.2022. DOI: [10.3390/jimaging8100285](#)
- Abhishek Dutta / Giles Bergel / Andrew Zisserman: Visual Analysis of Chapbooks Printed in Scotland. In: HIP '21: Proceedings of the 6th International Workshop on Historical Document Imaging and Processing (Lausanne, 06.09.2021). New York 2021. DOI: [10.1145/3476887.3476893](#)
- Hassan El-Hajj / Oliver Eberle / Anika Merklein / Anna Siebold / Noga Shlomi / Jochen Büttner / Julius Martinetz / Klaus-Robert Müller / Grégoire Montavon / Matteo Valleriani: Explainability and Transparency in the Realm of Digital Humanities: Toward a Historian XAI. In: International Journal of Digital Humanities 5 (2023), pp. 229–331. 02.10.2023. DOI: [10.1007/s42803-023-00070-1](#)
- Ute Frietsch: Alchemie-Notationen in IconClass. In: Alchemiegeschichtliche Quellen in der Herzog August Bibliothek. Wolfenbüttel 2017. HTML. [\[online\]](#)
- Jonathan Gaede: »So nehmet die materia und thut sie in ein solches Glas«. Gefäßdarstellungen in Destillationsbüchern und alchemistischen Traktaten der frühen Neuzeit. In: Bettina Lindner-Bornemann / Sebastian Kürschner (eds.): Die Sprache wissenschaftlicher Objekte. Interdisziplinäre Perspektiven auf die materielle Kultur in den Wissenschaften (= Lingua Academica, 8). Berlin etc. 2024, pp. 7–52. DOI: [10.1515/9783111437392-002](#)
- Germaine Götzelmann: Bilderschätze, Bildersuchen: Digitale Auswertung von Illustrationswiederverwendungen im Buchdruck des 16. Jahrhunderts. In: Philipp Hegel / Michael Krewet (eds.): Wissen und Buchgestalt (= Episteme in Bewegung, 26). Wiesbaden 2022, pp. 323–340. [\[Nachweis im GVK\]](#)
- Thomas Hofmeier: Woodcuts for Alchemists. Strategies of Illustrated Alchemical Books in Basel. In: Quaerendo 53 (2023), no. 3–4, pp. 198–232. [\[online\]](#)
- Sarah A. Lang / Bernhard Liebl / Manuel Burghardt: Toward a Computational Historiography of Alchemy: Challenges and Obstacles of Object Detection for Historical Illustrations of Mining, Metallurgy, and Distillation in 16th – 17th Century Print. In: Artjoms Šeļa / Fotis Jannidis / Iza Romanowska (eds.): Proceedings of the Computational Humanities Research Conference 2023 (CHR 2023, Paris, 06.–08.12.2023). Aachen 2023, pp. 29–48. PDF. [\[online\]](#)
- Sarah Lang: Alchemical Laboratories: Texts, Practices, Material Relics. An Introduction. In: Sarah Lang (ed.): Alchemische Labore. Texte, Praktiken und materielle Hinterlassenschaften / Alchemical Laboratories. Texts, Practices, Material Relics. Graz 2023, pp. 15–40. DOI: [10.25364/97839033740412](#)
- Sarah A. Lang (2024a): A Computer Vision Dataset for Alchemical Illustrations. GitHub. 2024. [\[online\]](#)
- Sarah Lang (2024b): How to Create Your Own Fine-Tuning or Training Dataset for Computer Vision Using Supervisely. In: The LaTeX Ninja Blog. 06.09.2024. [\[online\]](#)
- Stefan Laube: Am Anfang ist Gestaltung. Bemerkungen zu Titelblättern bei Destilliertraktaten des 16. Jahrhunderts. In: Philipp Hegel / Michael Krewet (eds.): Wissen und Buchgestalt (= Episteme in Bewegung, 26). Wiesbaden 2022, pp. 275–300. [\[Nachweis im GVK\]](#)
- Tsung-Yi Lin / Michael Maire / Serge Belongie / James Hays / Pietro Perona / Deva Ramanan / Piotr Dollár / Charles Lawrence Zitnick: Microsoft COCO: Common Objects in Context. In: David Fleet / Tomas Pajdla / Bernt Schiele / Tinne Tuytelaars (eds.): Computer Vision – ECCV 2014 (= Lecture Notes in Computer Science, 8693). Cham 2014, pp. 740–755. DOI: [10.1007/978-3-319-10602-1_48](#)
- Franklin F. Mendels: Proto-Industrialization: The First Phase of the Industrialization Process. In: The Journal of Economic History 32 (1972), no. 1, pp. 241–261. [\[online\]](#)
- Joseph Redmon / Santosh Divvala / Ross Girshick / Ali Farhadi: You Only Look Once: Unified, Real-Time Object Detection. ArXiv. 08.06.2015. DOI: [10.48550/arXiv.1506.02640](#)
- Ivo Purš / Vladimír Karpenko: The Alchemical Laboratory in Visual and Written Sources. Prague 2023. [\[Nachweis im GVK\]](#)
- Matteo Valleriani / Florian Kräutli / Daan Lockhorst / Noga Shlomi: Vision on Vision: Defining Similarities Among Early Modern Illustrations on Cosmology. In: Matteo Valleriani / Giulia Giannini / Enrico Giannetto (eds.): Scientific Visual Representations in History. Cham 2023, pp. 99–137. DOI: [10.1007/978-3-031-11317-8_4](#)
- Volkhard Wels: (Al)Chemisches Wissen im Buchdruck. In: Philipp Hegel / Michael Krewet (eds.): Wissen und Buchgestalt (= Episteme in Bewegung, 26). Wiesbaden 2022, pp. 159–126. [\[Nachweis im GVK\]](#)

List of Figures

Figure 1: Illustration of a furnace from Andreas Libavius' *Alchymia* (Frankfurt 1606), overlaid with segmentation and boundary box. [Source: Supervisely / WDB (Public Domain)]

Figure 2: Illustrations of various alchemical vessels from Johann Kunckel's *Der curieuses Kunst- und Werck-Schul Anderer Theil* (Nürnberg 1707, left) and Andreas Libavius' *Alchymia* (Frankfurt 1606, right). [Source: WDB, [oc-77-2](#) / [nd-4f-18](#) (Public Domain)]