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

Sarah A. Lang[1,*], Bernhard Liebl[2] and Manuel Burghardt[2]

[1]*Department Centre for Information Modelling (ZIM), University of Graz*
[2]*Computational Humanities Research Group, University of Leipzig*

Abstract
This study explores the use of modern computer vision methods for object detection in historical images extracted from 16th–17th century printed books containing illustrations of distillation, mining, metallurgy, and alchemical apparatus. We found that the transfer of knowledge from contemporary photographic data to historical etchings proves less effective than anticipated, revealing limitations in current methods like visual feature descriptors, pixel segmentation, representation learning, and object detection with YOLOv8. These findings highlight the stylistic disparities between modern images and early print illustrations, suggesting new research directions for historical image analysis.

Keywords
computer vision, object detection, alchemy, chymistry, early-modern print, metallurgy, mining, distillation, annotation

1. Illustrations in early modern technical literature

In the past few decades, the field of history of science has undergone a transformation, especially when it comes to the historiography of alchemy [46]. With the emergence of the so-called 'New Historiography of Alchemy' [45] and the rise of the *Experimental History of Science* [25, p. 85], traditional textual analysis has been supplemented by experimental methods often referred to as reconstruction, replication, and re-enactment, or the 'RRR methods' [29, p. 314]. As we move into the 2000s and beyond, the application of these methods has been widely recognized and is now established as a standard in the history of science [38]. Running parallel to this experimental shift is a growing interest in crafts knowledge, best exemplified by the work of Pamela Smith in the *Making and Knowing* Project [54]. The 'making and knowing' that was

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France
*Corresponding author.
sarah.lang@uni-graz.at (S. A. Lang); bernhard.liebl@uni-leipzig.de (B. Liebl);
burghardt@informatik.uni-leipzig.de (M. Burghardt)
https://sarahalang.com (S. A. Lang);
https://www.mathcs.uni-leipzig.de/ifi/forschung/computational-humanities/ (B. Liebl);
https://www.mathcs.uni-leipzig.de/personenprofil/mitarbeiter/juniorprof-dr-manuel-burghardt (M. Burghardt)
0000-0002-9145-7238 (S. A. Lang); 0000-0002-8593-400X (B. Liebl); 0000-0003-1354-9089 (M. Burghardt)

the norm in the kitchens and makeshift laboratories of the past has been seen by Smith as the precursor to the natural sciences and chemistry as we know them today [55, p. 292]. Morris argues that chemical laboratories in the modern sense emerged with the replacement of multi-purpose or make-shift spaces, which were not specifically designed for carrying out chemical operations, with professionalized work environments for performing chemical and metallurgical operations [41, p. 19–20]. He further states that this rise of chemical laboratories coincides with the boom of a genre of metallurgical technical treatises [40]. Empirical evidence of these first laboratories remains scarce, with only a handful of alchemical laboratories discovered so far [34]. This is where early modern handbooks on distillation, metallurgy and mining, rich with illustrations, become invaluable. These texts provide unmatched insight into the laboratories, processes, and practices in the *artes technicae* at the time, illustrating the underpinnings of the era's chemistry and technology. Despite their significance for the history of technology and the Chemical Humanities [43], these books remain relatively understudied to this day.

1.1. Depicting mining, metallurgy, and distillation

During the proto-industrial revolution, mining and metallurgy flourished, leading to the emergence of encyclopedic compendia of technological apparatus and processes. These include works such as Georgius Agricola's *De re metallica libri XII* [3], Vannoccio Biringucci's *De la pirotechnia Libri X* [6], Lazarus Ercker's *Aula subterranea* [23], and Giambattista della Porta's *De distillatione libri IX* [44]. Metallurgical technical treatises began to become a staple in the genre of didactic manuals and were frequently accompanied by technical illustrations. Beginning with smaller treatises, grander montanist works started appearing by the mid-16th century, such as Vannoccio Biringuccio's *De la Pirotechnia* (1540) [6] and Georg Agricola's *De Re Metallica* (1556) [2]. This knowledge, always accessible in books, as Michael Giesecke has emphasized, was so attractive because it replaced the exchange with experts and, thus, often made expensive and time-consuming journeys unnecessary [27]. Consequently, ease of finding relevant passages, through fitting illustrations or knowledge organization tools such as indices, was pivotal to their success. Besides metallurgical-focused works, distillation treatises also became popular in the 16th century [35]. Particularly influential was Hieronymus Brunschwig's *Liber der Arte Distillandi* (Straßburg 1512) [11] or Walther Hermann Ryff's Distillation Book (Frankfurt am Main 1545) [50]. Brunschwig's treatises have been published in a bewildering variety of versions, translations, and re-editions [35, p. 284–287].[1]

1.2. Research agenda and the case for automatic object detection

Since book illustration was expensive, early modern printers opportunistically reused illustrations from woodcuts and copper plates, thereby separating the images from their original contexts. Thus, illustrations would be commissioned for one specific publication, rendering lots of detail and providing an alternative communication medium for the message expressed in the text of that particular book, and then reused in other contexts where they fitted more

[1] The Strasbourg doctor and pharmacist first published his *Liber de Arte Distillandi De Simplicibus* in 1500. This is referred to by research as the 'small distillation book'. Twelve years later, the author followed up with a more voluminous *Liber de Arte Distillandi De Compositis*, known as the 'large distillation book' [11].

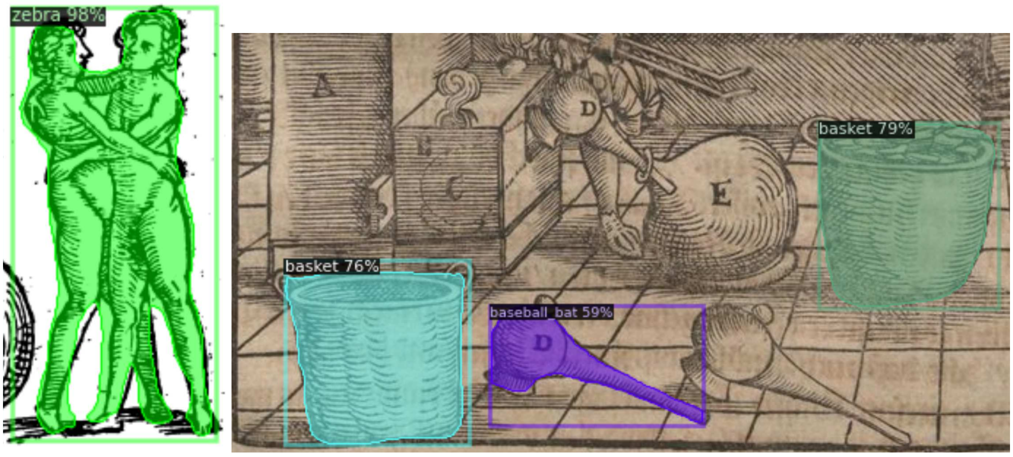


Figure 1: The Distant Viewing Toolkit [5] had some trouble labeling an allegorical depiction of an alembic [44, p. 42] as a zebra (green box, left). While some objects were detected correctly (blue boxes, right), there is an apparent labeling bias toward baseball bats [23, p. 166] (purple, right), as the model obviously has no appropriate labels for alchemical objects.

or less well, much like modern stock photography [28]. However, this means that not every image used in early modern print was made specifically to illustrate the exact matter discussed in a text passage. Medical books, herbiaries and distillation books are a medium particularly rich in illustration, for which even legal battles for 'copyright' are not unheard of. Especially later richly illustrated encyclopedic works could only be financed due to their reuse of earlier image material. What does this mean for pragmatic literature though? Do the images faithfully represent the processes being described and the equipment needed to carry them out? We know, for example, that Lazarus Ercker's *Aula Subterranea* [22] (or 'Bergwercksarten') is a true handbook, in the sense that it is detailed enough so that one can replicate the processes described. But can this be true for all other books from that genre as well, given what we know about the practices of illustration reuse in historical print?

It is in this context that we propose to apply computer vision techniques to automatically detect the illustrations in these books. Being able to detect relevant objects in digitised book pages is a crucial first step for a quantitative Distant Viewing [5] analysis of such apparatus within early modern chymical and pragmatic literature. In this short paper, we discuss challenges and obstacles we encountered during a first series of experiments in annotating a sample of such illustrations and training different approaches for object detection for historical illustrations of mining, metallurgy, and distillation in 16th–17th century print.

2. Detecting alchemical apparatus

2.1. Related work

We presume that a computational analysis of illustration practices can yield answers to the questions outlined above. As for related work, there is one branch of works that uses computer vision methods on illustrations in 15th/16th century print [37, 21, 28]. However, these approaches are less concerned with the recognition of individual objects and more focused on identifying illustrations as a whole, particularly their reuse in different books. Cormier et al. [17] use machine learning approaches to classify illustrations as either woodcut or copperplate engravings. An interactive Visual Analytics System (VeCHArt) for comparing copies or different states of a print is proposed by Pflüger et al. [42]. Valleriani et al. [58] present an empirical study on the visual similarity of early modern scientific illustrations on cosmology while Kaoua et al. [30] provide insights from a large-scale study on image collation, as they try to match different illustrations in a large corpus of manuscripts.

What all these approaches have in common is their emphasis on studying illustrations as complete entities, analyzing their style, similarities, or reuse. However, for our specific use case of detecting alchemical apparatus, we require an approach that is able to detect singular objects in a complex scene depicted in an illustration. Since we could not find any existing methods for object detection in 16th/17th century book illustrations, we conducted a series of experiments using various approaches on our own.

2.2. First experiments with existing methods

First of all, we experimented with out-of-the-box methods, such as the Distant Viewing Toolkit (Figure 1), Segment Anything (segment-anything.com/) (Figure 2) and image querying, using OWL-ViT (Figure 3). While revealing some successes at first glance, after some more testing these algorithms have proven largely inadequate in differentiating specific objects of interest from early modern prints. This medium is rich in visually similar etchings and contains typical alchemical objects that algorithms trained on modern data may simply not be familiar with.

2.3. Training and evaluation corpus

Because of the shortcomings of the above approaches, we proceeded to compile some training data, to provide a representative sample for the book genre defined above, containing books primarily concerned with mining, metallurgy or distillation. Some of them represent different issues or print runs of the same book for standard works such as Hieronymus Brunschwig's *De Arte Distillandi* [11], Georgius Agricola's *De re metallica* [3] or Vannoccio Biringuccio's *Pirotechnia* [6], in which illustrations frequently differ in between different editions or print shops. Unlike the training corpus, the evaluation corpus was constructed to contain books not only concerned with mining, metallurgy or distillation. This allows us to verify if the algorithm actually learned anything and is able to distinguish illustrations not related to our subject (such as workshop scenes not related to alchemy, metallurgy, mining or distillation) from the objects we wish to detect. Accordingly, we first evaluate the ability to detect illustrations of laboratory equipment in early modern book pages, and then look at the performance for classifying specific objects. Our training corpus, thus, only contains books that we know contain a sufficient number of relevant illustrations from the contexts of mining, metallurgy and distillation from the 16th-17th centuries [11, 12, 50, 51, 10, 13, 24, 57, 31, 20, 22, 3, 1, 2, 6, 7, 9, 8], while the evaluation corpus contains books from other alchemy-related areas and encyclopedias [56, 32,

Figure 2: A basic segmentation of alchemical apparatus worked out reasonably well with Segment Anything. However, this model is only capable of segmentation, but does not provide any labels for the segments.

33, 19, 18, 48, 4]. The challenges of annotating the training corpus are described in the next section.

3. The alchemy of annotation

The next step involved the semi-automatic annotation of images using the Supervisely platform (https://supervise.ly), whereby each component of alchemical apparatus was labeled individually in the hopes of providing the most useful form of annotations to improve model training. This process resulted in the creation of pixel-level labels.

We based our annotation on previous work done at the Herzog August Bibliothek Wolfenbüttel [26].[2] In Frietsch's classification [26], 'Alchemistic equipment' (49E393, https://iconclass.org/49E393) is a subclass of 'Alchemy' (49E39) in IconClass and organized as illustrated in Figure 4. As the annotation table (Table 1) shows, we did not incorporate all of the IconClass categories as labels. The classes to be used were selected by the relative frequency of related images in our corpus and depending on whether it made sense to keep subclasses or

[2]Adhering to the alchemy IconClass classification and vocabulary created by Ute Frietsch, which includes most alchemical apparatus, would not only keep a successful object detection model coming out of this work interoperable, but it also provides us with 1,800 tagged images we may re-use for creating ground truth in future work.

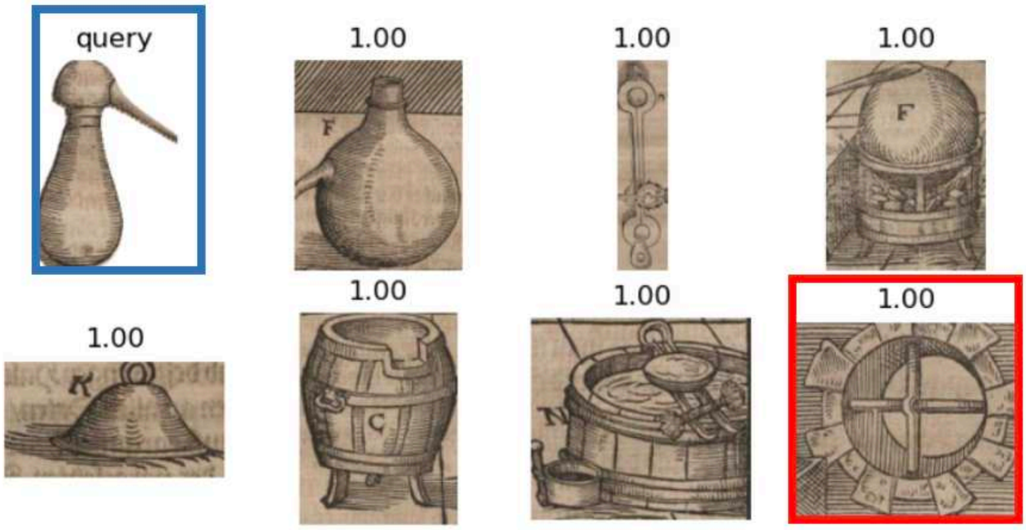


Figure 3: An example visual query (blue box) using OWL-ViT [39] on Lazarus Ercker's *Aula subterranea* [22, p. 74] illustrates that the model was able to detect all kinds of objects from workshop and laboratory contexts, but was not really capable of distinguishing alchemical from non-alchemical apparatus (for instance a window, see red box). The zero-shot approach OWL-ViT (Vision Transformer for Open-World Localization) is trained on image-text pairs to perform open-vocabulary object detection, allowing for image queries as well as text queries [39]. However, this method failed to generalise knowledge, possibly due to being trained primarily on modern, visually distinct objects, a stark contrast to the stylistically similar and overlapping nature of our early modern alchemical illustrations. When evaluating OWL-ViT, we only used image queries, which are transformed into CLIP by OWL-ViT. We could have tried to use the specialized terms as text queries in the hope that OWL-ViT might know them, but we did not attempt it because alchemical objects are hard to describe.

not (as many of them are not that visually distinctive nor frequent enough in our corpus to be effective to annotate). The goal was to keep the number of necessary annotation labels (and classes) as low as possible for our initial experiments. On the other hand, we introduced a class for *ambices* (singular *ambix*, a distillation helmet), which are frequently depicted, yet were lacking from Frietsch's classification of alchemical equipment.[3] This approach represents a compromise between keeping the number of classes as low as possible while still including a sufficient number for making meaningful interpretations later. Had we annotated both the non-explicitly alchemical and the explicitly alchemical tools the same way, we would probably train our algorithm to simply detect tools, regardless of the label assigned to them coming from the IconClass alchemy category.

[3]As we initially had planned not to include composite devices in the hopes of thus providing better training data for the algorithms, some classes very visually distinctive for alchemy were not included, such as alembics and moor's heads (Figure 5). Notably, within the category of 'pots' (*ollae*), some objects exhibit visually distinct alchemical characteristics, like triangular crucibles (for examples see Table 1), while others only can be interpreted as alchemical within a guaranteed alchemical context such as cupels, which visually look like simple pots or cups. We further opted to unite a range of furnace types under a single label.

49E3931 alchemistic vessels

- **49E39311** bottles (*ampullae*)
  - philosophical egg (*ovum philosophicum*)
  - pelican
  - phial
  - receiver (*receptaculum*)
- **49E39312** flasks (*cucurbitae*)
  - alembic
  - Moor's head
  - *operculum*
  - retort
  - rosenhut
- **49E39313** pots, jars (*ollae*)
  - aludel
  - chalice
  - crucible
  - cupel

49E3932 alchemistic furnace

- assay furnace
- athanor
- carburizing furnace
- 'slow Harry' (*piger henricus*)
- reverberatory furnace
- smelting furnace

49E3933 alchemistic bath (*balneum*)

- *balneum arenae*
- *balneum Mariae*

49E3939 other alchemistic equipment

Figure 4: The category 'Alchemistic equipment' (49E393, https://iconclass.org/49E393) from Ute Frietsch's IconClass tags for alchemy [26].

4. Prelimary results

In the rapidly evolving Digital Humanities (DH) sub-field of *Distant Viewing* [5], the application of computer vision techniques in diverse research areas has been met with enthusiasm. But despite this enthusiasm, our study reveals that these models may not yet readily adapt to specialized tasks in the DH. We have encountered substantial challenges in deploying these models for object detection in early modern depictions of chemical apparatus. The likely culprits were not solely the unique visual style of these etchings but also the models' unfamiliarity with the nuances of early modern alchemical equipment and associated terminology. It is apparent that these models, adept at interpreting modern visual styles and contexts, are confounded by the distinct visual style of early modern etchings. In the following subsections, we present preliminary results for the detection of alchemical objects in early modern illustrations that were achieved with a range of different supervised and unsupervised computer vision approaches.

Figure 5: Examples for classes not included (from left to right): alembik [11, p. 098] & moor's head [31, p. 8]. However, overlapping structures still appeared in the data anyway, so we may as well include those objects in the next attempt.

4.1. Visual feature descriptors

First, we experimented with an unsupervised clustering approach for visual feature description, namely the ORB (Oriented FAST and Rotated BRIEF [49]) method. This approach is tailored for exact image reproduction (cf. the work done on woodcut reuse in chapbooks with VISE [21]). This did not involve any training or the usage of our annotations and was meant to discern whether some intrinsic structures within the data could be utilized. Unfortunately, ORB failed to demonstrate such patterns in our data set.

4.2. Pixel Segmentation

Next, we decided to try pixel segmentation approaches, which allow us to perform object detection by dividing an image into segments and labeling each pixel, trying to map it to an object class. We first deployed approaches, where models classify each pixel individually, namely U-Net [47] (with a ResNet-34 backbone) and the newer SegFormer [59]. Despite being unable to recognize several elements (notably, animals), the U-Net/ResNet deep learning model detected, i.e. segmented, some plants correctly. Overall, however, the classification still proved to be erroneous. With the ResNet-based pixel segmentation, we reached an overall accuracy of 33.0% after fine tuning for 50 epochs. A similar story unfolded when using the SegFormer B1 [59] deep learning model, which occasionally managed to identify the rough area of an object but again without determining the correct category.

4.3. Representation learning

Furthermore, we continued assessing the efficacy of non-supervised models, which operate without annotation to discover structures in the data and thus are supposed to identify similar objects. We employed SimSiam (Simple Siamese Representation Learning) [16] and SimCLR (Siamese Contrastive Representation Learning) [15] for unsupervised clustering using Siamese networks. Siamese networks are used in unsupervised visual representation learning to maximize similarity between image augmentations. SimCLR (Contrastive Learning of Visual Representations) performs unsupervised representation learning from unlabeled images, which leverages data augmentation for contrasting different visual representations. SimCLR and SimSiam perform well on ImageNet. Yet the methods yielded equally discouraging results on our historical data.

4.4. YOLO object detection

Finally, we turned to the state-of-the-art object detection framework YOLO (*You Only Look Once*) version 8 model[4], because a predecessor (YOLOv5) had been previously reported as being suitable for detecting images in historical print [14]. Unfortunately, the performance of YOLO – like the previous approaches – fell short of our expectations. As YOLO is a popular framework and widely known in the Computational Humanities community, we will discuss it in more detail. We based our quality assessment on the model's ability to correctly detect objects and accurately label them.

YOLO training was performed using about 50% of each class for training and the rest for validation (figure 10). We initially experimented with 3-fold cross-validation, however, due to the scarceness of our training data, we finally opted for the single train-validation split.

As each image usually contained various labels with different classes (see Figure 6), producing such a stratified sampling was unfortunately not straightforward, as one image must be either assigned to the training set or to the validation set with all its containing labels to prevent data leakage. Sometimes a large proportion of all available labels was on one or two images. A further complication was the presence of sometimes partially overlapping label annotations, such as distillation helmets being part of furnace setups. These create potential sources of confusion for both training and validation (figure 6). We found no solution for fixing the overlapping labels, but we partitioned image regions with non-overlapping labels into isolated (non-overlapping and non-leaking) sub-images, thereby producing a larger number of possible assignments to training or validation sets (and through this a lower stratification error). The image partitioning was performed using a custom plane sweep algorithm that produced a hierarchy of either horizontal or vertical axes that subdivided images without cutting across label bounding boxes. To compute the actual training-validation split, we generated 10,000 random splits and picked the one that yielded a label distribution with the lowest mean error in its test-val ratio over all classes. For future studies, we plan to resort to more robust approaches [53]. Still, except for the furnace class, our approach produced a good stratification for all classes (figure 10).

[4]https://docs.ultralytics.com/

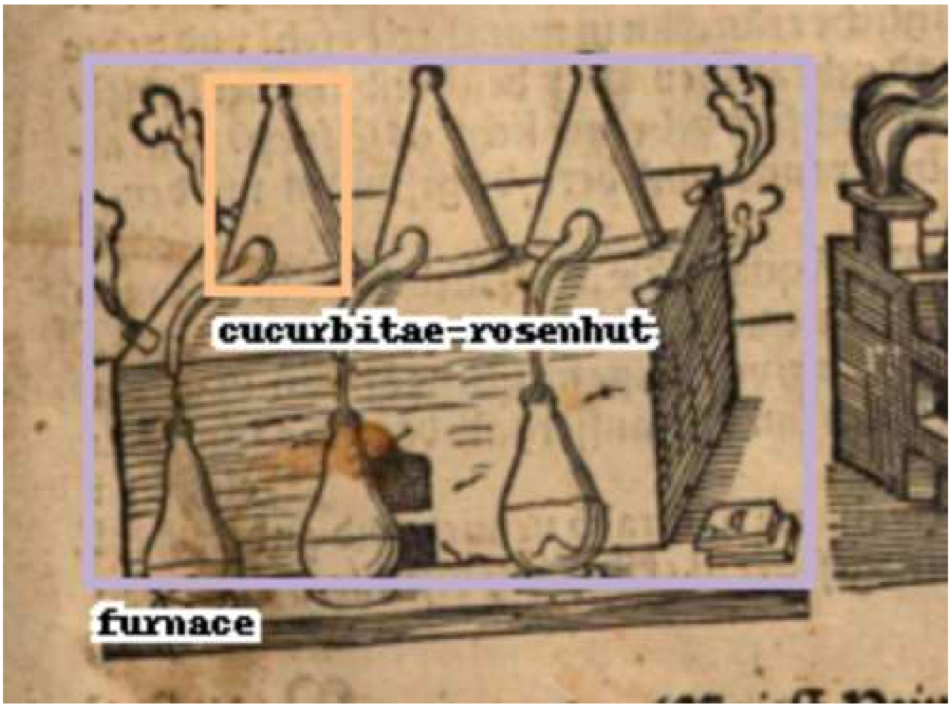


Figure 6: Example image illustrating annotation issues [31, p. 5]: a cucurbit is part of the composite furnace setup for an alchemical process, so we encounter both an overlapping annotation and, due to a number of decisions early on in the annotation process, only one of three of these cucurbits is annotated, potentially causing confusion for YOLO.

We now report some of the training results. Training a `yolo8n` model with default parameters yielded a model with a mAP@0.5 of 0.3. Switching to a `yolov8s` model with a resolution of 1280 pixels (instead of the 640-pixel default) improved this score to 0.37 (discussed below and shown in figure 7). As the confusion matrix (figure 8) and the precision-recall curves (figure 7) show, the classes that were best detected are 'plants', 'ollae' and 'animals'. 'Furnaces', 'other-equipment', 'cucurbitae-retorte', 'cucurbitae-rosenhut' and 'ampullae' are detected considerably less well, having both issues with precision and recall. The classes 'human', 'mineral-metal' and 'cucurbitae' showed very low overall precision. The detection of 'cucurbitae-ambix' did not seem to work at all. We also experimented with other resolutions (up to 1,600 pixels), as well as adding augmentation through `mixup` and various image transformations, as well as tuning the `mosaic` setting and the `box_loss` gain. However, we found no improvements in overall performance. Looking at the training curves, it turned out that for all tested YOLO models, resolutions, and settings, from smallest `yolov8n` model to the larger `yolov8s` model, generalization for object localization did not work well, while generalization for object classification seemed to present no issues at all: while the classification loss `cls_loss` was reduced rather symmetrically for both training and validation sets, and the `box_loss` for the training data showed a nearly perfect training curve in all regimes, the `box_loss` for the validation set turned out to be highly unstable and erratic in all cases, implying at least partial overfit-

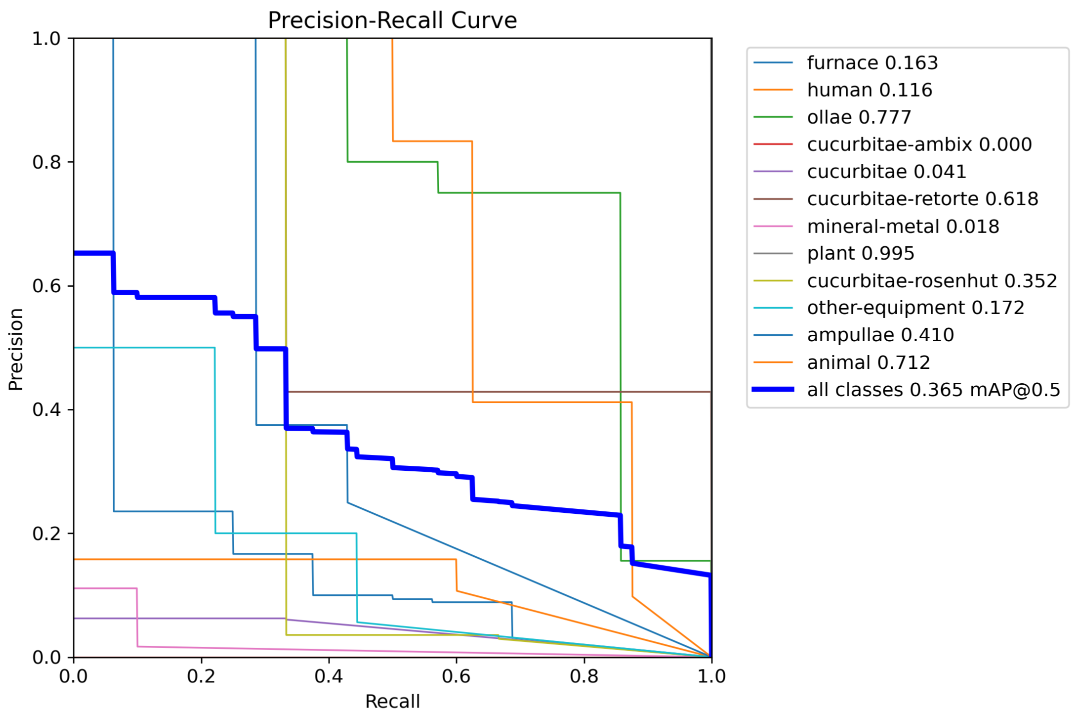


Figure 7: These precision-recall curves, markedly improved from our first attempts, show that the model is learning, albeit not equally well on every class.

ting. Upon analyzing why *ollae* was recognized better than other classes, we noted that the characteristic rounding could potentially account for a somewhat better model performance in this category. Across other label classes, visual variance was higher, which is illustrated in figure 10. For example, the depictions of objects in the *ampullae* category varied considerably (e.g., the jugs with handles in the 'training' set lacked larger openings at the top).

The overall lack of success was probably due to the ratio of large 'variance in the data' to small 'number of annotations'. The latter pales in comparison to the recommended figure of 1,500 images (and 10,000 labels) per category.[5] The daunting prospect of manually annotating such a volume of images, however, was contrary to our objectives of automating the task. Annotating 1,500 objects per category would not only be very laborious and potentially nonsensical for our task, this amount of examples per class also simply may not exist per object type in our historical data.

In preliminary experiments we observed that out-of-the-box YOLO models, pre-trained on COCO, showed no advantage in terms of transfer learning for the task at hand.[6] Not only are COCO images modern, but their differences also tend to be much bigger than amongst different types of early modern alchemical laboratory apparatus. Thus, the model probably cannot adapt

[5] https://docs.ultralytics.com/yolov5/tutorials/tips_for_best_training_results/
[6] The COCO dataset consists of 80 distinct object classes (from a modern context) like cats, zebras, or baseball bats.

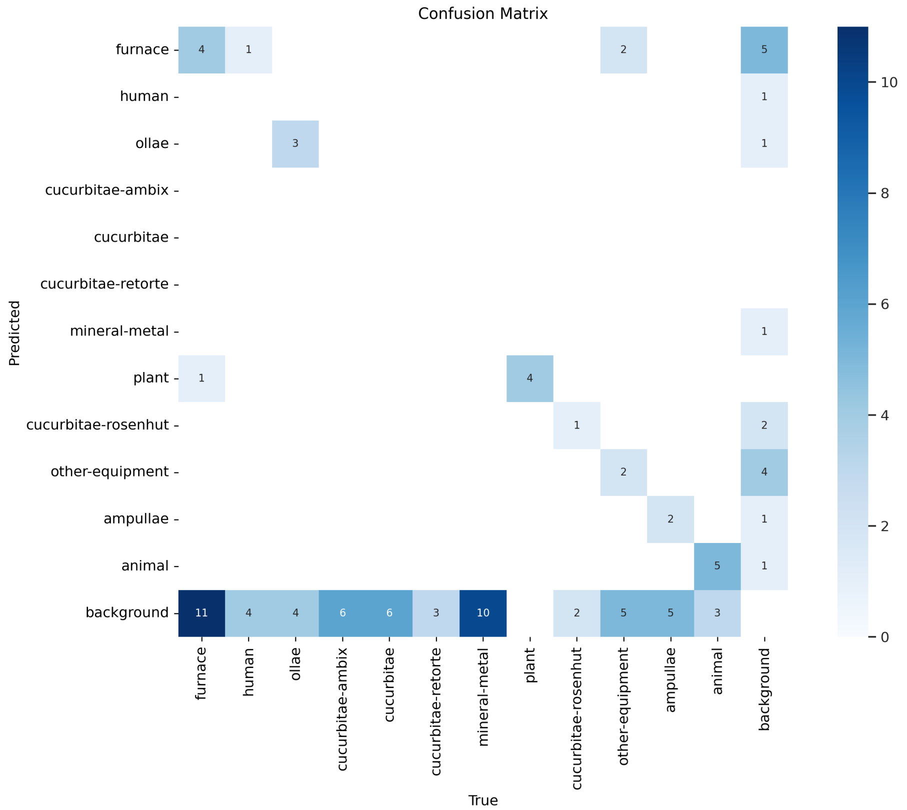


Figure 8: This confusion matrix indicates that the model learned something, even if the results remain poor. It seems that misclassification with other classes rarely happens, as nearly all confusions involve the background class, which might hint at a localization rather than a classification problem.

easily to grappling with historical data or distinguish in such a lot of details types of objects it has never seen before and does not know what to call.

5. Conclusions and future work

As part of our endeavour to utilize computer vision techniques for detecting early modern depictions of chemical apparatus, we initially embarked on experimental runs using readily available toolkits. These preliminary efforts yielded encouraging results, suggesting the viability of digging deeper into the intricacies of this interdisciplinary task. Encouraged by these early indications, we decided to extend our exploration, leveraging custom annotations to fine-tune a model. However, as the previous sections have detailed, these subsequent efforts were met with considerable obstacles and ultimately did not live up to the promise suggested by our initial forays. This, in turn, strongly suggests that further in-depth investigation is required in this area.

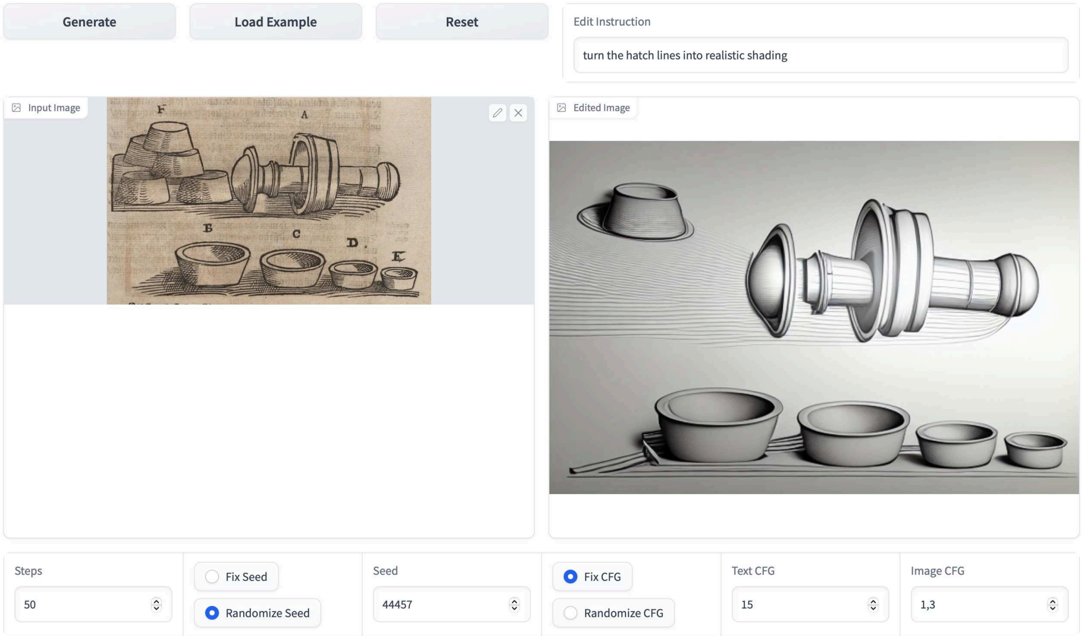


Figure 9: Attempting style transfer for early modern laboratory scenes [22, p. 12] using `instruct-pix2pix`. While this example looks promising, the algorithm actually lost visually distinctive features of alchemical objects in most cases, making this approach unsuitable for our purposes.

Attempts to harness state-of-the-art computer vision models revealed a distinct lack of generalisability to the idiosyncratic nature of early modern etchings. These unsuccessful attempts underscore the unique challenges presented by these unconventional early modern images. The 'rendering' or hatchings, i.e. the style of our images, could be what thwarts the algorithms. They may also have issues with granularity due to the etchings' visual similarity because all the objects to be analyzed are early modern book illustrations characterized by cross-hatching and strong black lines. The model may simply recognize them all as parts of books or book pages but does not realize that it is the difference between those particular illustrations that we are interested in.[7]

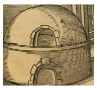
Going forward, we propose to explore one or few-shot approaches, although such methods are not extensively supported for object detection. We might try reducing our object detection problem (which is more complex than classification and for which there are also fewer readily available frameworks) into a classification problem by working with cropped images. The complex nature of the image data at hand suggests a need for more comprehensive annotation or potentially attempting to leverage style transfer to enhance our outcomes. We had initially tested this approach with the `InstructPix2Pix` model, which can convert hatching into real-

[7] This may be because the training data it was trained on probably did not have lots of images like ours and when it did, these simply may have been labelled as 'book' or 'book page' by annotators who, unlike us, were not interested in their particular details. At least indications for this were witnessed when we were first trying out the Distant Viewing Toolkit [5] as an out-of-the-box tool (cf. Figure 1).

istic shading, but unfortunately, this led to the loss of crucial visually distinctive details in the images and was ultimately unhelpful for our object detection task (Figure 9). Leveraging the classification capabilities of large Vision-Language Models (VLMs) such as BLIP-2 [36] would be very interesting as well, however, the object localization issue needs to solved first, maybe by using OWL-ViT [39] or SegmentAnything (Figure 2) only for bounding box estimation but not for classification [52].

In conclusion, despite the growing enthusiasm for *Distant Viewing* in the DH, the application of recent computer vision methods in the context of early modern print illustrations requires more nuanced approaches. The models' failure to recognize and classify early modern etchings of chemical apparatus serves as a sobering reminder of the gap that still exists between the out-of-the-box availability of state-of-the-art technology and the challenges in its DH application on historical data.

Table 1
Annotation Classes for Alchemical Objects

Label	Example Image	IconClass	Description	Visual Representation
human	 [22, p. 109]	3	Human Being, Man in General	Human figures, mostly men working in mines, but not exclusively. Their presence indicates a complex laboratory or mining scene.
animal	 [10, p. 125]	25F	Animals	Different species, sometimes present in distillation books.
plant	 [10, p. 101]	25G1	Plants (in general)	Illustrations of plants range from roots to all sorts of flowers. Most are visually quite similar to the non-botanist (presumably to the machine as well) and are present primarily in distillation books alongside distillation apparatus.
mineral-metal	 [3, p. 456]  [22, p. 238 (pdf)]	25D13	Minerals and Metals	Can be in the form of 'rocks' or ores but also molten.
furnace	 [1, p. 493]  [1, p. 437]	49 E 39 32	Alchemical furnaces	Alchemical furnaces, covering a whole range of sub-types (such as 'slow Harry' (*piger henricus*), athanor and assaying, carburizing, reverberatory, smelting or muffle furnaces), which can be quite visually different. Some are covered in fumes or fire.
cucurbitae	 [11, p. 048]	49 E 39 31 2	laboratory flasks (*cucurbitae*)	These flasks or flask-like vessels are hard to distinguish from *ampullae* and often do not contain many distinctive features on their own (except for when part of the special sub-classes below).
cucurbitae-ambix	 [11, p. 048 (pdf)]	not in IconClass	Distillation helmet	Included in IconClass only as part of alembics (which are actually composite devices); annotated with and without spout
cucurbitae-rosenhut	 [31, p. 015 (pdf)]	-	Special type of distillation helmet	Very distinctive triangular shape (triangular cylinder)
cucurbitae-retort	 [22, p. 143]	-	Particular type of flask (retort)	A flask combining *ambix* and *cucurbit*, which can be used instead of an alembic, and is very visually distinctive for alchemical equipment. It might make sense to include alembics in this class due to their similar shape and function.
ollae	 [22, p. 9]  [22, p. 24]	49 E 39 31 3	pots or jugs (*ollae*)	The class contains mostly crucibles (such as Hessian triangular crucibles) for pouring smolten metal or cupels (porous pots for cupellation, i.e. docimasy of metals, also called *testae*). Their depictions are too infrequent to further distinguish sub-classes.
ampullae	 [20, p. 39]	49 E 39 31 1	bottle-like containers (*ampullae*)	If the image does not show a particular type of bottle (such as the 'pelican' in the example image), distinguishing them from the flask category is non-trivial.
other-	 [8, p. 58]  [22,	49E3939	Other	Not strictly alchemical on its own

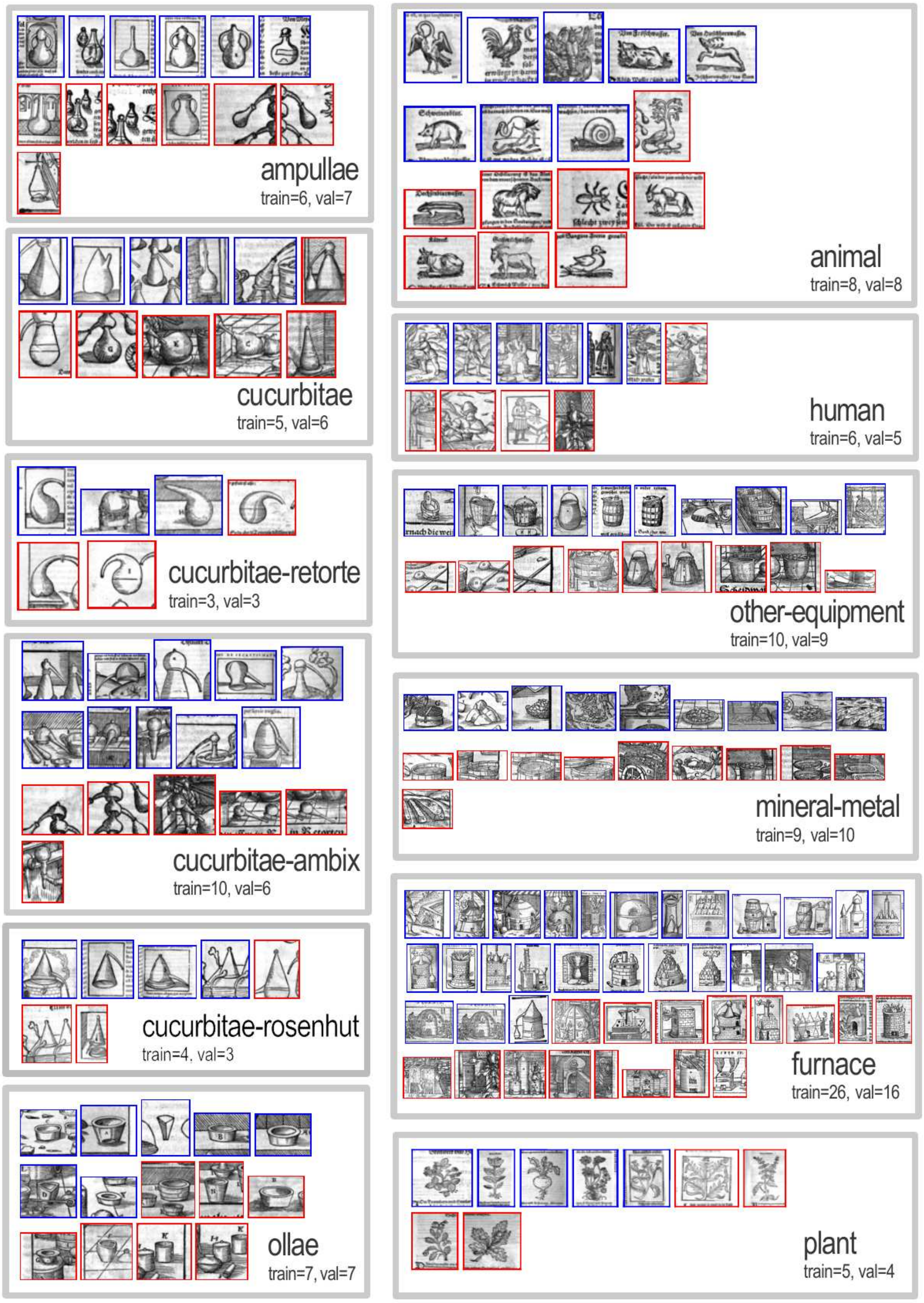


Figure 10: Training and validation sets, grouped by category. Items with blue borders are part of the training set, items with red borders are part of the validation set. The visual variation in most classes is quite high, especially those covering a general term such as 'animals' as opposed to, for example, cupels (*ollae*), which on top of the highly specialized nature of the concept are relatively simple structures visually.

References

[1] G. Agricola. *De Re Metallica*. Basileae / Basel: In Officina Frobeniana, Per Hier. Frobenivm Et Nic. Episcopivm, 1561. URL: http://data.onb.ac.at/rep/10B29699.

[2] G. Agricola. *De Re Metallica*. Basileæ / Basel: König, Emanuel, 1657. URL: http://data.onb.ac.at/rep/10B29680.

[3] G. Agricola. *Vom Bergkwerck*. Basel / Basel: Jeronymus Froben and Niclausen Bischoff, 1557. URL: http://data.onb.ac.at/rep/108C988B.

[4] U. Aldrovandi. *Musaeum Metallicum*. Bononiae / Bologna: Joan. Bapt. Ferronij, 1648. URL: http://data.onb.ac.at/rep/10B29761.

[5] T. B. Arnold and L. Tilton. "Distant Viewing: Analyzing Large Visual Corpora". In: *Digital Scholarship in the Humanities* (2019). DOI: http://dx.doi.org/10.1093/digitalsh/fqz013.

[6] V. Biringuccio. *Pirotechnia*. Vinegia / Venice: Giouan Padoano, 1550. URL: http://data.onb.ac.at/rep/108CB82E.

[7] V. Biringuccio. *Pirotechnia*. Venetia / Venice: Gironimo Giglio, 1559. URL: http://data.onb.ac.at/rep/1085DDED.

[8] V. Biringuccio. *Pirotechnia*. Vinegia / Venice: Comin da Trino di Monferrato : Navo, Curzio Troiano de, 1559. URL: http://data.onb.ac.at/rep/10727526.

[9] V. Biringuccio. *Pirotechnia*. Bologna / Bologna: Longhi, 1678. URL: http://data.onb.ac.at/rep/106F01F6.

[10] H. Brunschwig. *Distilierbuch der rechten Kunst*. Franckfurt am Mayn / Frankfurt am Main: Weygand Hanen Erben, 1565. URL: http://data.onb.ac.at/rep/1084C642.

[11] H. Brunschwig. *Liber de arte distillandi*. Straßburg / Strasbourg: Johann Grüninger, 1512. URL: http://data.onb.ac.at/rep/109559D1.

[12] H. Brunschwig. *The vertuose boke of distyllacyon of the waters of all maner of herbes*. London: Laurens Andrewe, 1527. URL: https://archive.org/details/mobot31753000816063/page/n30/mode/thumb.

[13] H. Brunschwig and W. H. Ryff. *New Vollkommen Distillierbuch*. Franckfurt a. M. / Frankfurt am Main: Christian Egenolffs Erben, 1597. URL: http://data.onb.ac.at/rep/10A5871B.

[14] J. Büttner, J. Martinetz, H. El-Hajj, and M. Valleriani. "CorDeep and the Sacrobosco Dataset: Detection of Visual Elements in Historical Documents". In: *Journal of Imaging* 8.10 (Oct. 15, 2022), p. 285. DOI: 10.3390/jimaging8100285.

[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. *A Simple Framework for Contrastive Learning of Visual Representations*. June 30, 2020. URL: http://arxiv.org/abs/2002.05709.

[16] X. Chen and K. He. *Exploring Simple Siamese Representation Learning*. Nov. 20, 2020. URL: http://arxiv.org/abs/2011.10566.

[17] M. Cormier, S. Park, and L. Beck. "Automatic Classification of Woodcuts and Copperplate Engravings". In: *2020 17th Conference on Computer and Robot Vision (CRV)* (2020), pp. 85–92. URL: https://api.semanticscholar.org/CorpusID:219548487.

[18] J. v. Cuba. *Gart der Gesuntheit*. Straßburg / Strasbourg: Mathia Apiario, 1536. URL: http://data.onb.ac.at/rep/108C96EA.

[19] J. v. Cuba. *Ortus Sanitatis*. Venetiis / Venice: Joannes de Cereto de Tridino, 1538. URL: http://data.onb.ac.at/rep/109E61DA.

[20] G. Della Porta. *De Distillatione*. Romae / Rome: Reu. Cameræ Apostolicæ, 1608. URL: http://data.onb.ac.at/rep/1058CD41.

[21] A. Dutta, G. Bergel, and A. Zisserman. "Visual Analysis of Chapbooks Printed in Scotland". In: *The 6th International Workshop on Historical Document Imaging and Processing (HIP '21), September 5–6, 2021, Lausanne, Switzerland* (2021). DOI: https://doi.org/10.1145/3476887.3476893.

[22] L. Ercker. *Aula Subterranea*. Franckfurt am Mayn / Frankfurt am Main: Johan Feyerabendt, 1598. URL: http://data.onb.ac.at/rep/104A2CBD.

[23] L. Ercker. *Aula subterranea*. Frankfurt (Main): Zunner, 1672. URL: https://www.deutschestextarchiv.de/book/show/ercker%5C%5Faula01%5C%5F1672.

[24] M. Ficinus. *Liber de arte distulandi*. Straßburg / Strasbourg: Johan Grüniger, 1509. URL: http://data.onb.ac.at/rep/10A81A42.

[25] H. Fors, L. M. Principe, and H. O. Sibum. "From the Library to the Laboratory and Back Again: Experiment as a Tool for Historians of Science". In: *Ambix* 63/2 (2016), pp. 85–97.

[26] U. Frietsch. *IconClass alchemy notation*. Herzog August Bibliothek Wolfenbüttel, 2017. URL: http://alchemie.hab.de/bilder.

[27] M. Giesecke. *Der Buchdruck in der frühen Neuzeit. Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien*. Frankfurt am Main, 1991.

[28] G. Götzelmann. "Bilderschätze, Bildersuchen: Digitale Auswertung von Illustrationswiederverwendungen im Buchdruck des 16. Jahrhunderts". In: *Wissen und Buchgestalt*. Ed. by P. Hegel and M. Krewet. Wiesbaden: Harrassowitz Verlag, 2022, pp. 323–340.

[29] M. M. A. Hendriksen. "Rethinking Performative Methods in the History of Science". In: *Berichte zur Wissenschaftsgeschichte* 43 (2020), pp. 313–322.

[30] R. Kaoua, X. Shen, A. Durr, S. Lazaris, D. Picard, and M. Aubry. "Image Collation: Matching illustrations in manuscripts". In: *CoRR* abs/2108.08109 (2021). URL: https://arxiv.org/abs/2108.08109.

[31] P. Kerzenmacher. *Alle Farben Wasser*. München / Munich, Frankfurt am Mayn / Frankfurt am Main: Egenolff, 1589. URL: https://digital.deutsches-museum.de/de/digital-catalogue/library-object/BV020856683/%5C#12.

[32] A. Kircher. *Mundus subterraneus 1*. Amstelodami / Amsterdam: Janssonius & Weyerstraten, 1665. URL: http://data.onb.ac.at/rep/1047DA01.

[33] A. Kircher. *Mundus subterraneus 2*. Amstelodami / Amsterdam: Janssonius & Weyerstraten, 1665. URL: http://data.onb.ac.at/rep/1047D9FF.

[34] S. Lang. “Alchemical Laboratories: Texts, Practices, Material Relics. An Introduction”. In: *Alchemistische Labore. Praktiken, Texte und materielle Hinterlassenschaften / Alchemical Laboratories. Practices, texts, material relics.* Ed. by S. Lang, M. Fröstl, and P. Fiska. Graz: Grazer Universitätsverlag, 2023, pp. 13–30.

[35] S. Laube. “Am Anfang ist Gestaltung. Bemerkungen zu Titelblättern bei Destilliertraktaten des 16. Jahrhunderts”. In: *Wissen und Buchgestalt.* Ed. by P. Hegel and M. Krewet. Wiesbaden: Harrassowitz Verlag, 2022, pp. 275–298.

[36] J. Li et al. “BLIP-2: Bootstrapping Language-Image Pre-Training with Frozen Image Encoders and Large Language Models”. In: *arXiv preprint arXiv:2301.12597* (2023). URL: http://arxiv.org/abs/2301.12597.

[37] M. Malaspina and Y. Zhong. “Image-matching technology applied to Fifteenth-century printed book illustration”. In: *Lett Mat Int* 5 (2017), pp. 287–292.

[38] M. Martinón-Torres. “Some recent developments in the historiography of alchemy”. In: *Ambix* 58/3 (2011), pp. 215–37.

[39] M. Minderer, A. Gritsenko, A. Stone, M. Neumann, D. Weissenborn, A. Dosovitskiy, A. Mahendran, A. Arnab, M. Dehghani, Z. Shen, X. Wang, X. Zhai, T. Kipf, and N. Houlsby. “Simple Open-Vocabulary Object Detection with Vision Transformers”. In: *Eccv 2022.* 2022. DOI: https://doi.org/10.48550/arXiv.2205.06230.

[40] P. J. T. Morris. “The history of chemical laboratories: a thematic approach”. In: *ChemTexts* 7/21 (2021). DOI: https://doi.org/10.1007/s40828-021-00146-x.

[41] P. J. T. Morris. *The Matter Factory: A History of the Chemistry Laboratory.* London: Reaktion Books, 2015.

[42] H. Pflüger, D. Thom, A. Schütz, D. Bohde, and T. Ertl. “VeCHArt: Visually Enhanced Comparison of Historic Art Using an Automated Line-Based Synchronization Technique”. In: *IEEE Transactions on Visualization and Computer Graphics* 26.10 (2020), pp. 3063–3076. DOI: 10.1109/tvcg.2019.2908166.

[43] M. Piorko, M. Hendriksen, and S. Werrett. “Alchemical Practice: Looking Towards the Chemical Humanities”. In: *Ambix* 69:1 (2022), pp. 1–18. DOI: 10.1080/00026980.2022.2035572.

[44] G. D. Porta. *De Distillationibus libri IX.* Strasbourg: Zetzner, 1609.

[45] L. M. Principe and W. R. Newman. “Some Problems with the Historiography of Alchemy”. In: *Secrets of Nature: Astrology and Alchemy in Early Modern Europe.* Ed. by W. R. Newman and A. Grafton. Cambridge/Massachusetts: MIT Press, 2001, pp. 385–432.

[46] S. Reardon. “The Alchemical Revolution”. In: *Science* 332 (2011), pp. 914–915.

[47] O. Ronneberger, P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation.* May 18, 2015. URL: http://arxiv.org/abs/1505.04597.

[48] E. Rößlin and J. v. Cuba. *Kreuterbuch.* Franckenfurt, am Meyn / Frankfurt am Main: Egenolff, 1536. URL: http://data.onb.ac.at/rep/109E61B8.

[49] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. “ORB: An efficient alternative to SIFT or SURF”. In: *2011 International Conference on Computer Vision*. 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: Ieee, Nov. 2011, pp. 2564–2571. DOI: 10.1109/iccv.2011.6126544.

[50] W. H. Ryff. *Das new groß Distillier Buch*. Franckfurt / Frankfurt am Main: Christ. Egenolff, 1545. URL: http://data.onb.ac.at/rep/107BF132.

[51] W. H. Ryff. *New groß distillier-Buch*. Franckfort / Frankfurt am Main: Christian Egenolff’s Erben, 1556. URL: http://data.onb.ac.at/rep/105FA8A2.

[52] P. Saranrittichai et al. “Zero-Shot Visual Classification with Guided Cropping”. In: *arXiv preprint arXiv:2309.06581* (2023). URL: http://arxiv.org/abs/2309.06581.

[53] K. Sechidis et al. “On the Stratification of Multi-Label Data”. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by D. Gunopulos et al. Vol. 6913. Springer Berlin Heidelberg, 2011, pp. 145–158. DOI: https://doi.org/10.1007/978-3-642-23808-6_10.

[54] P. Smith. *The Making and Knowing Project (2015–2022)*. 2020. URL: makingandknowing.org.

[55] P. H. Smith. “Laboratories”. In: *The Cambridge History of Science 3 / Early Modern Science*. Ed. by K. Park and L. Daston. Cambridge: Cambridge University Press, 2006, pp. 290–305.

[56] L. Thurneysser zum Thurn. *Historia sive Descriptio Plantarum Omnium*. Berlini / Berlin: ?, 1578. URL: http://data.onb.ac.at/rep/108C96EA.

[57] P. Ulsted and J. A. et al. *Coelum philosophorum*. Lugduni / Lyon: Guilielmus Rovillius, 1572. URL: http://data.onb.ac.at/rep/1086E9DB.

[58] M. Valleriani, F. Kräutli, D. Lockhorst, and N. Shlomi. “Vision on Vision: Defining Similarities Among Early Modern Illustrations on Cosmology”. In: *Scientific Visual Representations in History*. Ed. by M. Valleriani, G. Giannini, and E. Giannetto. Cham: Springer International Publishing, 2023, pp. 99–137. DOI: https://doi.org/10.1007/978-3-031-11317-8_4.

[59] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*. Oct. 28, 2021. URL: http://arxiv.org/abs/2105.15203.